AD-609378

AFSES-
64-11
c.1

LOAN COPY: RETURN TO
AFWL (WLIL-2)
KIRTLAND AFB, N MEX

11th ANNUAL AIR FORCE SCIENCE AND ENGINEERING SYMPOSIUM

20-22 OCTOBER 1964

HOSTS

AEROSPACE MEDICAL DIVISION
AIR FORCE SYSTEMS COMMAND
OFFICE OF AEROSPACE RESEARCH

BROOKS AIR FORCE BASE, TEXAS

Chairman
DR. WINSTON MARKEY, CHIEF SCIENTIST
UNITED STATES AIR FORCE

20080215309

HEADQUARTERS
AEROSPACE MEDICAL DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
BROOKS AIR FORCE BASE, TEXAS
78235

20 October 1964

SUBJECT: 11th Annual Air Force Science and Engineering Symposium

TO: Participants of the 11th Annual Air Force Science and Engineering Symposium

1. It is a great pleasure for me, as Commander of the Aerospace Medical Division, and on behalf of the Air Force Systems Command and the Office of Aerospace Research, to welcome you to the 11th Annual Science and Engineering Symposium. The aerospace growth of our nation as a whole has been remarkable in the two years since we last hosted the Science and Engineering Symposium here at Brooks.

2. The topics to be discussed at this meeting will afford a close look at several diverse aspects of this growth. From this symposium will come a better understanding of our aerospace capability and limitations.

T. C. Bedwell, Jr.

T. C. BEDWELL, JR.
Major General, USAF, MC
Commander

PREFACE

The Air Force Science and Engineering Symposium is held to demonstrate the capability of the Air Force and the competence of its scientists and engineers to conduct and manage high quality in-house research programs. It is particularly intended to emphasize the overall in-house capabilities of the Air Force Systems Command (AFSC) and the Office of Aerospace Research (OAR), and to demonstrate the effectiveness of the Air Force laboratories.

The symposium program presents the spectrum of science and engineering activity as carried out by AFSC and OAR. This activity includes research, exploratory development, engineering development, advanced development, test and evaluation. Toward this end, the program consists of a selection of the best "contributed" papers submitted in competition for the Aerospace Education Foundation Award. The scientific and technical papers to be presented will provide a representative sample of the scope and depth of Air Force research, development, technology and engineering, and of its value to the national defense.

The symposium program is presented before a distinguished and critical audience assembled from the Department of Defense and other government scientific agencies.

INDEX OF PAPERS

Presented at 11th Annual Air Force Science and Engineering Symposium
Aerospace Medical Division
Brooks Air Force Base, Texas
20-22 October 1964

1. SUITABLE LUNAR LANDING SYSTEM CONSIDERING THE BEARING CAPACITY AND FAILURE MODES OF LUNAR SURFACE MATERIALS
   Lt Colonel John D. Peters
   Arnold Engineering Development Center
   Air Force Systems Command

2. MINIMUM FUEL TRAJECTORIES WITH SOFT LANDINGS
   Major Orlando J. Manci, Jr.
   Frank J. Seiler Research Laboratory
   Office of Aerospace Research

3. NEW SELECTIVELY ABSORBING COATINGS FOR SOLAR POWERED THERMOELECTRIC DEVICES
   1st Lt Ronald L. Long
   Air Force Materials Laboratory
   Research and Technology Division
   Air Force Systems Command

4. ENGINEERING SUPPORT FOR ELECTRIC PROPULSION SPACE TESTS
   Capt Peter D. Tannen
   Air Force Special Weapons Center
   Air Force Systems Command

5. CARDIOVASCULAR RESPONSES TO GRAVITATIONAL CHANGES AFTER PROLONGED BED REST
   Major Perry B. Miller
   Lt Colonel Robert L. Johnson
   Lawrence E. Lamb, M. D.
   USAF School of Aerospace Medicine
   Aerospace Medical Division
   Air Force Systems Command

6. INTERDISCIPLINARY MEASUREMENT OF HUMAN PERFORMANCE UNDER LOW AND ZERO GRAVITY CONDITIONS
   Dr. Duane F. Kasten
   Aerospace Medical Research Laboratories
   Aerospace Medical Division
   Air Force Systems Command

7. THE EFFECTS OF SECOBARBITAL AND D-AMPHETAMINE ON PERFORMANCE DURING A SIMULATED TACTICAL AIR MISSION

Major Richard E. McKenzie
Lois L. Elliott, Ph.D.
USAF School of Aerospace Medicine
Aerospace Medical Division
Air Force Systems Command

8. DIRECT ELECTROFLUIDDYNAMIC ENERGY CONVERSION PROCESSES FOR POWER GENERATION

M. O. Lawson
H. J. P. von Ohain
Aerospace Research Laboratories
Office of Aerospace Research

9. ZERO GRAVITY ELECTROLYSIS CELL

1st Lt Michael D. Rutkowski
1st Lt Michael A. Maxwell
Air Force Flight Dynamics Laboratory
Research and Technology Division
Air Force Systems Command

10. INVESTIGATION OF SEMICONDUCTOR EQUALIZER DIODES FOR SECONDARY BATTERY CELLS

Wayne S. Bishop
Air Force Aero Propulsion Laboratory
Research and Technology Division
Air Force Systems Command

11. HARMONIC GENERATION OF MICROWAVE PHONONS

Paul H. Carr
Air Force Cambridge Research Laboratories
Office of Aerospace Research

12. SPECTRA OF ANIONS: ACETYLACETONATES

Carl A. Pitha
Air Force Cambridge Research Laboratories
Office of Aerospace Research

13. MEASUREMENTS OF RADIO STAR AND SATELLITE SCINTILLATIONS AT A SUB-AURORAL LATITUDE

Richard S. Allen
Jules Aarons
Herbert Whitney
Air Force Cambridge Research Laboratories
Office of Aerospace Research

14. ELECTRON RADIATION DAMAGE IN ZnSe AT LOW TEMPERATURE

B. A. Kulp
Capt R. M. Detweiler
Aerospace Research Laboratories
Office of Aerospace Research

15. MODIFICATION OF THE EFFECTS OF ACUTE LETHAL DOSES OF IONIZING RADIATION IN THE RHESUS PRIMATE

Capt George S. Melville, Jr.
George W. Harrison, Jr.
Major Robert J. Young
USAF School of Aerospace Medicine
Aerospace Medical Division
Air Force Systems Command

16. APPLICATION OF ENZYMES IN PHOTOCHROMIC OPTICAL SHUTTERS FOR NUCLEAR FLASH PROTECTION

Sidney Allinikov
Air Force Materials Laboratory
Research and Technology Division
Air Force Systems Command

17. A STUDY OF PANEL THICKNESS REQUIRED TO PREVENT FLUTTER IN ADVANCED PERFORMANCE VEHICLES

Capt Bobby L. Langley
Air Force Flight Dynamics Laboratory
Research and Technology Division
Air Force Systems Command

18. ANALYTICAL INVESTIGATION OF EFFICIENCY IN STRUCTURAL VIBRATION WITH AND WITHOUT DAMPING TREATMENT

Channing L. Pao
Systems Engineering Group
Research and Technology Division
Air Force Systems Command

19. CRACK PROPAGATION IN CUMULATIVE DAMAGE FATIGUE TESTS

Harry A. Lipsitt
Aerospace Research Laboratories
Office of Aerospace Research

20. THE STUDY OF AN IGNITION TECHNIQUE FOR LARGE SOLID ROCKET MOTORS

Major Martin McDermott Bretting
1st Lt Walter R. Niessen
Air Force Rocket Propulsion Laboratory
Research and Technology Division
Air Force Systems Command

21. A RANGE SAFETY SYSTEM FOR MISSILES WITH TOXIC PROPELLANTS

D. A. Haugen
Air Force Cambridge Research Laboratories
Office of Aerospace Research

22. A NEW CONCEPT IN THRUST BEARINGS FOR OSCILLATORY MOTION

E. J. Thomas
Rodney L. Clark
Systems Engineering Group
Research and Technology Division
Air Force Systems Command

23. SPACE USES OF THE EARTH'S MAGNETIC FIELD

1st Lt Ralph B. Hoffman
Lt Col Thomas O. Haig
Space Systems Division
Air Force Systems Command

24. A METHOD FOR CONTROLLING THE ATTITUDE OF A SATELLITE USING THE EARTH'S MAGNETIC FIELD

Wilfred J. Minkus
Rome Air Development Center
Research and Technology Division
Air Force Systems Command

25. LIGHTNING - AN AEROSPACE NEMESIS?

Major James H. Scharff
Air Force Weapons Laboratory
Research and Technology Division
Air Force Systems Command

26. USE OF INFRARED INSTRUMENTATION FOR DETECTION OF JET ENGINE TURBINE WHEEL TEMPERATURE

Fred L. Young
Air Force Logistics Command

27. DETERMINATION OF STABILITY DERIVATIVES FROM FLIGHT TEST DATA BY MEANS OF OPERATION ANALOG MATCHING

1st Lt Johnny M. Rampy
Donald T. Berry
Air Force Flight Test Center
Air Force Systems Command

28. RADIOISOTOPE TECHNIQUES FOR THE MEASUREMENT OF VTOL LANDING AND TAKEOFF PARAMETERS

Paul Polishuk
1st Lt Rudy C. Beavin
Air Force Flight Dynamics Laboratory
Research and Technology Division
Air Force Systems Command

SRL 64-3

Frank J. Seiler Research Laboratory

30 May 1964

Minimum Fuel Trajectories with Soft Landings

Maj Orlando J. Manci, Jr.

Contents

List of Symbols	i
List of Figures	iii
Abstract	iv
Introduction	1
Text	3
Conclusion	20
Figures	22
References	36
Appendix I	38
Acknowledgement	42

List of Symbols (in order of their definition)

F	=	thrust
x	=	distance in horizontal direction
y	=	altitude in vertical direction
t	=	time
m	=	mass of system
g	=	local gravity acceleration
R	=	distance traveled in horizontal direction during flight
θ	=	angle of rocket thrust with respect to the horizontal
α	=	angle of the tangent to the trajectory with respect to the horizontal
c	=	rocket exhaust velocity
β	=	mass-burning rate
ϕ_β	=	normalized mass-burning rate function
ϕ_s	=	switching function
J	=	payoff functional
$\vec{p}$	=	adjoint vector
H	=	Pontryagin's maximum function
p	=	magnitude of second and fourth components of the adjoint vector
θ_{min}	=	minimum angle of the thrust vector that will give a vertical acceleration equal to the local gravity
t_{co}	=	time at which rocket cuts off
t_{ri}	=	time at which rocket reignites
$(\bar{\ })$	=	physical quantity with dimensions (time, distance, or mass)
$(\ )_o$	=	initial condition, e.g., $t_o = 0$

$( )_1$ = terminal condition, e.g., $x_1 = R$

$( )^i$ = $i^{\underline{th}}$ component of a vector, e.g., p^4

$( )_{co}$ = condition at rocket cutoff

$( )_{ri}$ = condition at rocket reignition

List of Figures

Fig.	Title	Page
1	Problem Coordinates $(\bar{x}, \bar{y})$ Domain	22
2	Analog Computer Schematic	23
3	$\dot{x}$, $\dot{y}$, and y vs. time	24
4a	Sample Trajectories $m_o = 0.4$	25
4b	Thrust Angle vs. Time $m_o = 0.4$	25
5	p_o^1 vs. Range (R)	26
6	p_o^2 and p_o^3 vs. Range (R)	27
7	Rocket Cut-off Time (t_{co}) vs. Range (R)	28
8	Rocket Reignition Time (t_{ri}) vs. Range (R)	29
9	Time of Landing (t_1) vs. Range (R)	30
10	Initial Thrust Angle (θ_o) vs. Range (R)	31
11	Initial Thrust Angle (θ_o) vs. Initial Mass (m_o)	32
12	Angles of Thrust and Trajectory Tangents vs. Initial Mass at $t = 0$, t_{co}, t_{ri}, and t_1	33
13	Final to Initial Mass Ratio $\frac{m_1}{m_o}$ vs. Range	34
14	Horizontal Velocity $(\dot{x}_{co})$ at Cut-off vs. Range (R)	35

Abstract

Minimum fuel trajectories for rocket flights with soft landings in a uniform gravitational field are presented as functions of two parameters. The solutions are valid for any uniform gravitational field, rocket fuel, and burning rate.

The problem is formulated as a flight over a flat moon without atmosphere. The solution is subject to the differential constraints of the equations of motion and the end point conditions which require zero velocity at take-off and landing. This system of equations involving five parameters is reduced to a two-parameter system by means of dimensional analysis. Pontryagin's maximum principle is employed to reduce the problem to a two-point boundary-value problem with the optimality condition included. The resultant formulation requires the determination of four undertermined multipliers.

This two point boundary value problem is solved on the electronic differential analyzer. The solutions are presented as: (1) curves of the initial condition of each Lagrange multiplier, (2) rocket cutoff and reignition times, (3) total flight time, (4) initial thrust angle, and (5) final mass fraction, all as functions of range. The initial mass is the parameter in each of these families of curves.

A method for mechanizing this solution suitable for one man "back pack" type as well as heavier systems is described. This open loop control system requires only a simple stable platform and cam follower driven by a constant speed motor to achieve the desired trajectories.

These results can be used as a standard for evaluation of other trajectories and modes of transportation as well as directly for translation of men and equipment over specified distances.

Introduction

The problem of traveling from one point to another on a flat earth has long been a practical consideration for man. With the prospect of manned exploration of planets and satellites having little or no atmosphere to impede or disturb flight, the use of rocket propulsion, a long-standing dream, becomes a very attractive and practical means of moving men and equipment over unprepared terrain. As the aerospace mission of the Air Force developes, it will be necessary to understand the advantages offered by the several means of locomotion and to have quantitative information upon which to base a choice. The minimum fuel trajectories with soft landing presented in this paper provide such information for travel through negligible atmospheres over distances such that the gravity field can be considered uniform, e.g., ranges of 30 miles and altitudes of 5 miles on the moon. The solutions can be used directly as a basis for the design of a "back-pack" system as suggested below; as a standard against which the cost, in terms of fuel, of other systems can be compared; or, as a planning instrument in deploying satellite bases about a central base when fuel conservation is a paramount consideration, as it must be when fuel is to be supplied from the earth.

The problem posed and solved in this paper is that of a mass which starts from rest, is propelled a given distance by means of a rocket over a trajectory which requires minimum fuel and returns to zero altitude with zero velocity. The equations of motion are derived in dimensional form and are reduced to dimensionless equations by means of a suitable set of dimensionless ratios. In this form, the minimum fuel requirement is incorporated into the equations and the problem reduced to a two point boundary value problem by the application of Pontryagin's maximum principle. The resulting equations are solved on an analog computer to produce solutions which are valid for all rocket fuels, all burning rates, all gravity fields, and encompass all possible initial thrust accelerations from impulsive thrust to the minimum thrust for which the solutions exist. These solutions are presented as functions of range with initial mass as a parameter and as functions of initial mass with range as a parameter.

Statement of the Problem

The equations of motion for a particle with mass, $\bar{m}$, moving in a uniform gravitational field, g; and being acted upon by a thrust, $\bar{F}$; can be readily obtained by the application of Newton's second law to the force diagram of Figure 1. These equations of motion, valid at any instant of time, the initial and final conditions of the problem are given in equations (1)

$$\frac{d^2\bar{x}}{d\bar{t}^2} = \frac{\bar{F}}{\bar{m}}\cos\theta$$

$$\frac{d^2\bar{y}}{d\bar{t}^2} = \frac{\bar{F}}{\bar{m}}\sin\theta - g \qquad (1)$$

$$\bar{x}(0) = \bar{x}_o = 0 \qquad \frac{d\bar{x}}{d\bar{t}}(0) = 0$$

$$\bar{y}(0) = \bar{y}_o = 0 \qquad \frac{d\bar{y}}{d\bar{t}}(0) = 0$$

$$\bar{x}(t_1) = \bar{x}_1 = \bar{R} \qquad \frac{d\bar{x}}{d\bar{t}}(\bar{t}_1) = 0$$

$$\bar{y}(t_1) = \bar{y}_1 = 0 \qquad \frac{d\bar{y}}{d\bar{t}}(t_1) = 0$$

The angle, θ, and the thrust magnitude, $\bar{F}$, are the control variables in the problem. A thrust program which will provide for translation over a range, $\bar{R}$, landing with the prescribed terminal conditions and using the minimum amount of fuel

possible must be determined. The necessary conditions for such a minimum fuel trajectory will be obtained by applying Pontryagin's maximum principle to obtain additional equations which will incorporate this minimum fuel requirement into the equations of motion. The resulting equations will then form a two point boundary value problem whose solution is the desired trajectory. Before proceeding to this point, however, the characteristics of the rocket thrust, including a maximum burning rate limit, will be introduced and the parameters of the problem reduced to the minimum possible by dimensional analysis.

The thrust, $\bar{F}$, of a rocket may be expressed as the product of the rocket exhaust velocity, c, and the mass burning rate, β, [1 p 377]

$$\bar{F} = c\beta \tag{2}$$

It is assumed that the thrust is limited by the burning rate so that

$$0 \leq \beta \leq \beta_{max} \tag{3}$$

The mass is then a function of time with a rate of change equal to the negative of the burning rate. Thus

$$\frac{d\bar{m}}{d\bar{t}} = -\beta \tag{4}$$

Substitution of equations (2), (3), and (4) into equations (1) yields equations (5)

$$\frac{d^2\bar{x}}{d\bar{t}^2} = \frac{c\beta}{\bar{m}(\bar{t})}\cos\theta$$

$$\frac{d^2\bar{y}}{d\bar{t}^2} = \frac{c\beta}{\bar{m}(\bar{t})}\sin\theta - g$$

$$\frac{d\bar{m}}{d\bar{t}} = -\beta$$

$$\bar{m}(0) = \bar{m}_o \qquad (5)$$

$$\bar{x}(0) = \bar{y}(0) = \frac{d\bar{x}}{d\bar{t}}(0) = \frac{d\bar{y}}{d\bar{t}}(0) = 0$$

$$\bar{x}(\bar{t}_1) = \bar{R}$$

$$\bar{y}(\bar{t}_1) = \frac{d\bar{x}}{d\bar{t}}(\bar{t}_1) = \frac{d\bar{y}}{d\bar{t}}(\bar{t}_1) = 0$$

Non-dimensional form

In order to eliminate the dimensions and reduce the problem to a minimum number of parameters, equations (5) are transformed by equations (6), derived by means of Buckingham's Pi Theorem [2], to obtain equations (7)

$$x = \frac{g}{c^2}\bar{x}$$

$$y = \frac{g}{c^2}\bar{y} \qquad (6)$$

$$t = \frac{g}{c}\bar{t} \qquad (6)$$

(cont)

$$m = \frac{g}{c\beta_{max}}$$

$$\phi_\beta = \frac{\beta}{\beta_{max}} \qquad 0 \leq \phi_\beta \leq 1$$

The dimensionless distance variables are actually ratios of the potential energy of a mass at a point $\bar{x}$ or $\bar{y}$ in a gravitational field g to the kinetic energy this mass would have if it were moving at the exhaust gas velocity c.

The new time variable is the ratio of velocity a unit mass would attain if accelerated in a gravity field, g, during a time, $\bar{t}$, to the rocket exhaust velocity, c.

The dimensionless mass is the ratio of the weight of the system in a gravitational field, g, to the magnitude of the thrust. The mass variable is then the reciprocal of the thrust to weight ratio. Intuitively initial thrust to weight ratios of from one to infinity should be of interest. However, it will be shown that minimum fuel trajectories exist only for m_o, the reciprocal of the initial thrust to weight ratio, in the range zero to approximately 0.62, if negative values of y are excluded from the trajectories.

When equations (6) are substituted into equations (5) and the variables u and v are defined as $\frac{dx}{dt}$ and $\frac{dy}{dt}$

respectively; the set of five simultaneous first order differential equations (7) results. Here differentiation with respect to t is indicated by the dot notation.

$$\begin{aligned}
\dot{x} &= u & x_o &= 0;\ x_1 = R \\
\dot{u} &= \frac{\phi_\beta}{m}\cos\theta & u_o &= 0;\ u_1 = 0 \\
\dot{y} &= v & y_o &= 0;\ y_1 = 0 \\
\dot{v} &= \frac{\phi_\beta}{m}\sin\theta - 1 & v_o &= 0;\ v_1 = 0 \\
\dot{m} &= -\phi_\beta & m_o &= m_o
\end{aligned} \tag{7}$$

Pontryagin's maximum principle will now be applied to equations (7), to obtain the necessary conditions for the minimum fuel trajectories.

Pontryagin Maximum Principle and Analysis

The development in this paper follows directly from the description of the maximum principle given in [3]. Kopp has obtained essentially the same equations in [4] using Rozonoer's formulation [5]. Miele, [4], obtains an equivalent set of equations by means of the classical calculus of variations with Valentine's extension [6].

The pay-off functional which is to be minimized is defined as

$$J = \int_o^{t_1} \phi_\beta dt \tag{8}$$

This functional is, of course, the mass of the fuel used. From this definition a sixth state variable x^o is defined[1] as

$$\dot{x}^o = \phi_\beta \qquad x^o_o = 0 \tag{9}$$

The state vector is thus

$$x = \begin{bmatrix} x^o \\ x^1 \\ x^2 \\ x^3 \\ x^4 \\ x^5 \end{bmatrix} = \begin{bmatrix} x^o \\ x \\ u \\ y \\ v \\ m \end{bmatrix} \tag{10}$$

The vector p is defined to be

$$p = \begin{bmatrix} p^o \\ p^1 \\ p^2 \\ p^3 \\ p^4 \\ p^5 \end{bmatrix} \tag{11}$$

such that

$$\frac{dp^i}{dt} = - \sum_{j=0}^{5} \frac{\partial f^j}{\partial x^i} \, p^j$$

[1]From this point superscripts will be used to indicate components of vectors and the subscript $( )_o$ to indicate values of state or adjoint variables at time $t = 0$, and $( )_1$ at time $t = t_1$.

where $\dot{x}^j = f^j$.

The maximum function is formed as the inner product of equations (7) augmented by (9) with equation (11). Thus, if H denotes the maximum function

$$H = p^0\phi_\beta + p^1 u + p^2\left(\frac{\phi_\beta}{m}\cos\theta\right) + p^3 v + p^4\left(\frac{\phi_\beta}{m}\sin\theta - 1\right) - p^5\phi_\beta \qquad (12)$$

From Theorem 1, Chapter I, of [3], this function must take on its maximum value throughout the trajectory for an optimum choice of controls ϕ_β and θ . Further, since H is not an explicit function of time, this maximum is in fact zero at all times [3] . Using this fact the thrust program is readily obtained.

The control space may be considered to be the surface of a right circular cylinder with unit radius in the $\sin\theta$, $\cos\theta$ plane and extending from this plane to a height of unity in the ϕ_β direction. The angle θ may vary over all possible values so

$$\frac{\partial H}{\partial \theta} = 0$$

is a necessary condition for an extremum and (13)

$$\frac{\partial^2 H}{\partial \theta^2} < 0$$

when evaluated at the extremum is a sufficient condition for a

maximum in θ . Evaluation of these two partical derivatives yields the condition

$$\tan\theta = \frac{p^4}{p^2} \tag{14}$$

for optimal thrust direction control. The use of the sufficient condition resolves the sign ambiguity which results from the necessary condition [4].

The possible values of ϕ_β are limited by a saturation constraint on the burning rate. The evaluation of the maximum function with respect to this variable is thus somewhat different. In this case, the contribution of the terms in ϕ_β is made maximum at all times. This condition may be obtained by considering the terms containing ϕ_β in equation (12)

$$\phi_\beta \left(\frac{p^2}{m} \cos\theta + \frac{p^4}{m} \sin\theta + p^0 - p^5 \right)$$

The contribution of these terms to the maximum function, H , will be a maximum if ϕ_β takes on its maximum value of unity when the term in parentheses is positive and its minimum value of zero when this term is negative. Thus the rocket must burn at its maximum when it burns and be cut off at all other times. The term in parentheses determines when the thrust is switched on and off and is defined to be the switching function, ϕ_s .

$$\phi_s = p^0 + \frac{p^2}{m} \cos\theta + \frac{p^4}{m} \sin\theta - p^5 \tag{15}$$

This equation may be simplified by defining p to be

$$p = \sqrt{(p^2)^2 + (p^4)^2} \quad . \tag{16}$$

This definition leads to $\cos\theta = \frac{p^2}{p}$ and $\sin\theta = \frac{p^4}{p}$. Substitution of equations (16) into (15) yields

$$\phi_s = p^o + \frac{p}{m} - p^5 \tag{17}$$

The thrust program is thus seen to be a function of the $\vec{p}$ vector and the mass.

The differential equations of the components of $\vec{p}$ are obtained by the relationships

$$\frac{dp^i}{dt} = -\frac{\partial H}{\partial x^i} \tag{18}$$

as shown in [3]. Equation (18) applied to equation (12) yields

$$\begin{aligned} \dot{p}^o &= 0 \\ \dot{p}^1 &= 0 \\ \dot{p}^2 &= -p^1 \\ \dot{p}^3 &= 0 \\ \dot{p}^4 &= -p^3 \\ \dot{p}^5 &= \frac{\phi}{m^2}\beta p \end{aligned} \tag{19}$$

Equations (19) for p^1, p^2, p^3, and p^4 can be integrated to obtain

$$\begin{aligned} p^1 &= p_o{}^1 \\ p^2 &= p_o{}^2 - p_o{}^1 t \\ p^3 &= p_o{}^3 \\ p^4 &= p_o{}^4 - p_o{}^3 t \end{aligned} \tag{20}$$

Instead of solving for p^5 from equations (19), it is more convenient to obtain the differential equation for ϕ_s, with its initial condition to determine the switching time because ϕ_s incorporates the information contained in the equations for p^o and p^5, determines the switching times directly, and can be evaluated at initial and terminal conditions. Thus, from equations (17) and evaluation of the maximum function at $t = 0$,

$$\begin{aligned} \dot{\phi}_s &= \frac{\dot{p}}{m} \\ \phi_s(0) &= \frac{{p_o}^4}{\phi_\beta} \\ \dot{p} &= -({p_o}^1 \cos\theta + {p_o}^3 \sin\theta) \end{aligned} \tag{21}$$

It is easily shown that $\dot{\phi}_s$ can change sign at most once, so ϕ_s can change sign at most twice. Thus, in order to launch above the plane, it is necessary and sufficient that, $\phi_s(0) > 0$, $\dot{\phi}_s(0) < 0$, and $\dot{v}(0) > 0$. The trajectory must consist of a thrust arc, coast arc, and thrust arc, in that order.

An additional check on the solution is obtained by evaluating equation (12) at time $t = t_1$ to obtain

$$\phi_s(t_1) = \frac{{p_1}^4}{\phi_\beta(t_1)} \ . \tag{22}$$

Thus, the final set of equations with initial and terminal conditions which describe the minimum fuel trajectories, are given as equations (23).

$$
\begin{aligned}
\dot{x} &= u & x_o &= 0 & x_1 &= R \\
\dot{u} &= \frac{\phi_\beta}{m}\cos\theta & u_o &= 0 & u_1 &= 0 \\
\dot{y} &= v & y_o &= 0 & y_1 &= 0 \\
\dot{v} &= \frac{\phi_\beta}{m}\sin\theta - 1 & v_o &= 0 & v_1 &= 0 \\
\dot{m} &= -\phi_\beta & m(0) &= m_o & & \\
\dot{\phi}_s &= \frac{\dot{p}}{m} & \phi_s(0) &= \frac{p_o^4}{\phi_\beta(0)} & &
\end{aligned}
\qquad (23)
$$

$$\dot{p} = -(p_o^1 \cos\theta + p_o^3 \sin\theta) \qquad p_o = \sqrt{(p_o^2)^2 + (p_o^4)^2}$$

$$p^2 = p_o^2 - p_o^1 t$$

$$p^4 = p_o^4 - p_o^3 t$$

In order to solve equations (23) and satisfy both initial and final conditions, the proper values of p_o^1, p_o^2, p_o^3, and p_o^4 must be obtained. The problem may be simplified by noting that the equations for p^2 and p^4 are homogeneous in the p's. Thus, p_o^4 is chosen to be unity and the other initial values of the adjoint variables specified as ratios of p_o^4.

Solution of the equations

Equations (23) can be solved by either analog or digital computers. The method of solution must provide for:

(1) estimating the initial values of the adjoint variables,

(2) integrating the equations to t_1^* which is chosen as the time at which $\dot{x}$ returns to zero,

(3) determining the errors in y and $\dot{y}$, and

(4) changing the initial values of the adjoint variables to correct these errors.

The analog computer permits a man to be an integral part of the computation loop and has proven to be successful in the solution of equations (23).

The solution of equations (23) requires an open loop integration to obtain the state variables. Thus, extreme care must be exercised in scaling the equations, eliminating sources of noise and drift, and minimizing the time of solution. This last factor, however, must be balanced with the requirement that switching times and transients must be negligible because the solution proved to be extremely sensitive to delays in switching the function ϕ_β between its limits. The solutions discussed below were obtained by means of the analog computer circuit shown in Figure 2 (which is labeled with unscaled variables) on an Electronics Associates 231R computer. They were spot checked by a digital computer solution using the MIDAS Program [7] on an IBM 7094 computer. The analog and digital results agreed to three significant figures indicating that the open loop integration is sufficiently accurate.

Results

Initial experimentation with the equations showed that:

(1) p_o^1 has the greatest effect on the range;

(2) p_o^2 affects the initial slope of $\phi_s(t)$, the minimum value of ϕ_s and t_{co} and t_{ri}, the times at which ϕ_s becomes 0;

(3) p_o^3 affects essentially only t_{ri}.

Thus, the solutions were obtained by selecting values of p_o^1 and varying p_o^2 and p_o^3 to obtain the required terminal conditions.

A representative solution is shown in Figure 3 in which $\dot{x}$, $\dot{y}$, and y are plotted against a common time axis. The solution is indicated because these three curves reach the time axis at the same instant. This time is fortunately independent of amplitude scaling. Figure 4 shows two sample trajectories and the corresponding thrust angle programs.

The values of the initial conditions of the adjoint variables which produce solutions are plotted in Figures 5 and 6 as functions of range with initial mass as a parameter. These curves can be used to obtain the initial values of the Lagrange multipliers for any given range and, with interpolation between curves, for any initial mass. These values can then be used in equations (14) and (20) to obtain the required rocket nozzle angle program as a function of time.

The rocket burning program for any range is obtained from the curves of Figures 7, 8, and 9, which show the cutoff time, t_{co}, reignition time, t_{ri}, and landing time, t_1, as functions of range with initial mass as a parameter. These curves are also plotted for an initial mass of zero

which corresponds to the pure ballistic trajectory with impulsive thrust at launch and at landing.[2] The impulsive thrust case corresponds to an initial mass of zero because the non-dimensional mass is the reciprocal of the thrust to weight ratio.

The initial thrust acceleration determines the minimum initial angle, θ_{min}, of the thrust that will give a positive vertical acceleration since

$$\dot{v}_o = \frac{\phi_\beta(0)}{m_o} \sin\theta_o - 1$$

If $\dot{v}_o = 0$ then $\sin\theta_{min} = m_o$, from equations (23). Thus, θ_{min} increases as m_o, the reciprocal of the initial thrust to weight ratio, increases. However, Figure 10 shows that θ_o decreases as m_o increases in order to obtain the minimum fuel trajectory. This fact is emphasized in Figure 11 where θ_o, a weak function of range, is plotted against m_o, for two values of range. From this figure, it is readily seen that there is a maximum initial mass at each range for which the minimum fuel trajectories exist because $\dot{v}_o < 0$ below the boundary $\sin\theta_o = m_o$. The intersection of the θ_o vs m_o curves with this boundary marks the maximum value of m_o for each range. The solutions shown thus cover the entire permissible range of m_o for the horizontal range studied.

[2]See Appendix I for the derivation of the equations for impulsive thrust.

Figure 12 shows the thrust vector angles, θ, at t_o, t_{co}, t_{ri}, and t_1 and the tangents to the trajectories at t_{co} and t_{ri}, as functions of m_o. In each case these angles are weak functions of range and approach the well known optimum launch angle of 45° for impulsive thrust $m_o = 0$ for all ranges. The tangent angles at cutoff and reignition, α_{co} and α_{ri}, show that the trajectories become flatter as m_o increases. This fact is corroborated by the ratio of maximum altitude to range, which varies smoothly from $\frac{1}{4}$ for the ballistic trajectory to .01 for $m_o = 0.6$, and the curves of Figure 13 which show $\dot{x}$ at cutoff to increase for a given range as m_o increases.

The fuel required for each trajectory is shown in Figure 14 where the ratio of final to initial mass, $\frac{m_1}{m_o}$ is shown as a function of range with initial mass as a parameter. The impulsive thrust program is seen to be most efficient with more fuel being required as the initial thrust acceleration is decreased.

Mechanization of the control system

The curves of Figures 5 through 9 provide a complete set of data for the determination of the rocket thrust direction program, ignition, cutoff and reignition times, and final cutoff. These data are sufficient for a simple open loop control system which requires only a reference

platform for orientation of the rocket thrust and a timing system for control of the thrust vector. The timing could be obtained from a simple cam driven by a constant speed motor. The cam would contain thrust direction, or rocket nozzle position, as a continuous function of time with stops for t_{co}, t_{ri}, and t_1. Each cam would be designed for a specific range and initial mass so a cam-changing arrangement would permit rapid, easy selection of the desired trajectory.

It is emphasized here that this study has considered the motion of a point mass only. The dynamics of a distributed mass system would have to be included before the system could be mechanized.

Experience with the analog computer has shown that it is extremely difficult to obtain exactly a zero velocity on landing. This difficulty is caused in large part by the sensitivity of the terminal velocity to the reignition time, t_{ri}, and maximum errors on the order of 10^{-2} of the maximum y velocity and position can be expected. This amounts, for example, to a maximum error of approximately 52 feet in altitude and 1 ft/sec in terminal y velocity for a range of 28,500 ft on the moon with $m_o = 0.4$.[3]

[3]This initial mass corresponds to an initial weight of 1600 earth pounds if the burning rate is 2 earth pounds per second and a fuel with $I_{sp} = 320$ sec on earth is used.

Thus, a shock absorbing system and a backup rocket system for landing or main rocket failure should be included in any serious design. In a back-pack application the man could absorb landing shocks if he can correct terminal errors with such a backup system.

Conclusion

The solution of the minimum fuel trajectory problem with soft landing presented here has several direct appplications, viz., the design of rocket propulsion systems for distances up to 100 miles on the moon, a standard for comparison of other possible trajectories or systems of locomotion to determine their relative cost in terms of fuel required, a planning instrument for location of depots and satellite bases on extraterrestrial bodies. In their general nondimensional form they are applicable to any gravitational field, rocket fuel or burning rate. While not in closed form [4], the graphical form amplified by graphical techniques permits their use for the most general conditions.

Further specific solutions can always be calculated very simply with only a nominal amount of analog computing equipment. Solutions for soft landings at altitudes different from the launch altitude can also be easily obtained.

In the expanding mission of the Air Force, in which it will be necessary to explore and establish bases in unfamiliar, unimproved areas, rocket propulsion will be among the first means of locomotion between points for men and materials. The solutions presented here will aid in the design of systems

[4]The closed form is not likely to be found due to the nonlinearity of the problem.

which will guarantee the most economical use of valuable fuel. The simple control system suggested here can be used for unmanned transportation of materiel or transportation of personnel not skilled in flying, thereby making possible quick exploitation of new terrain and facilitating exploration.

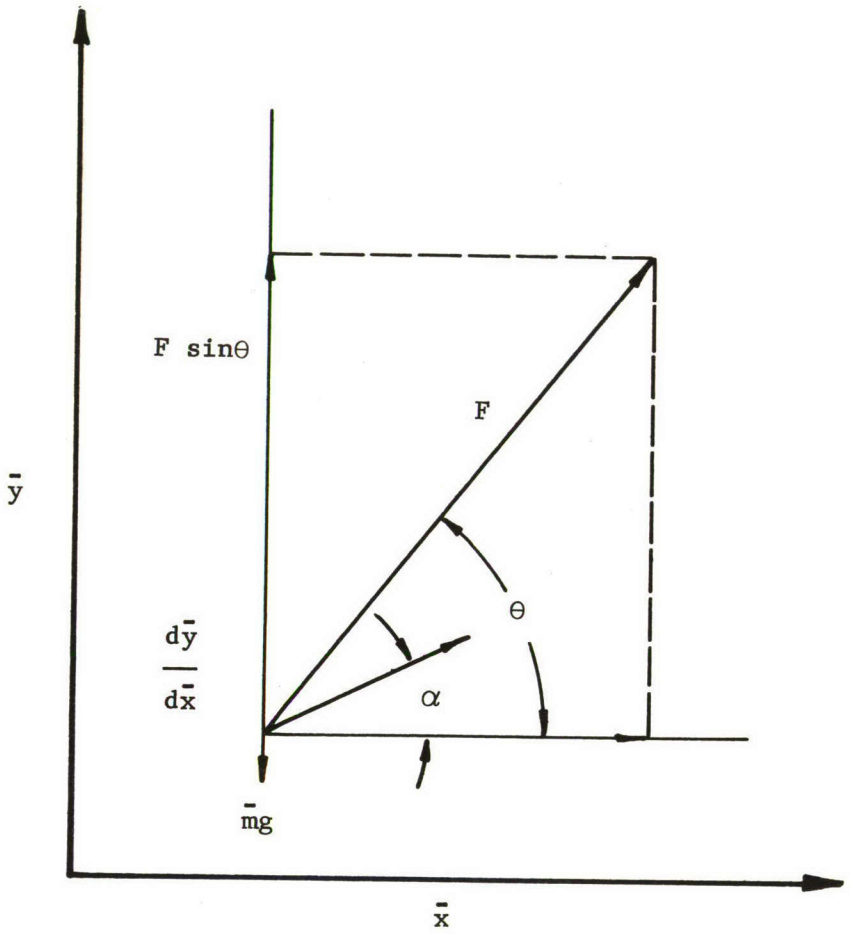


Fig. 1 Problem Coordinates
$(\bar{x},\bar{y})$ Domain

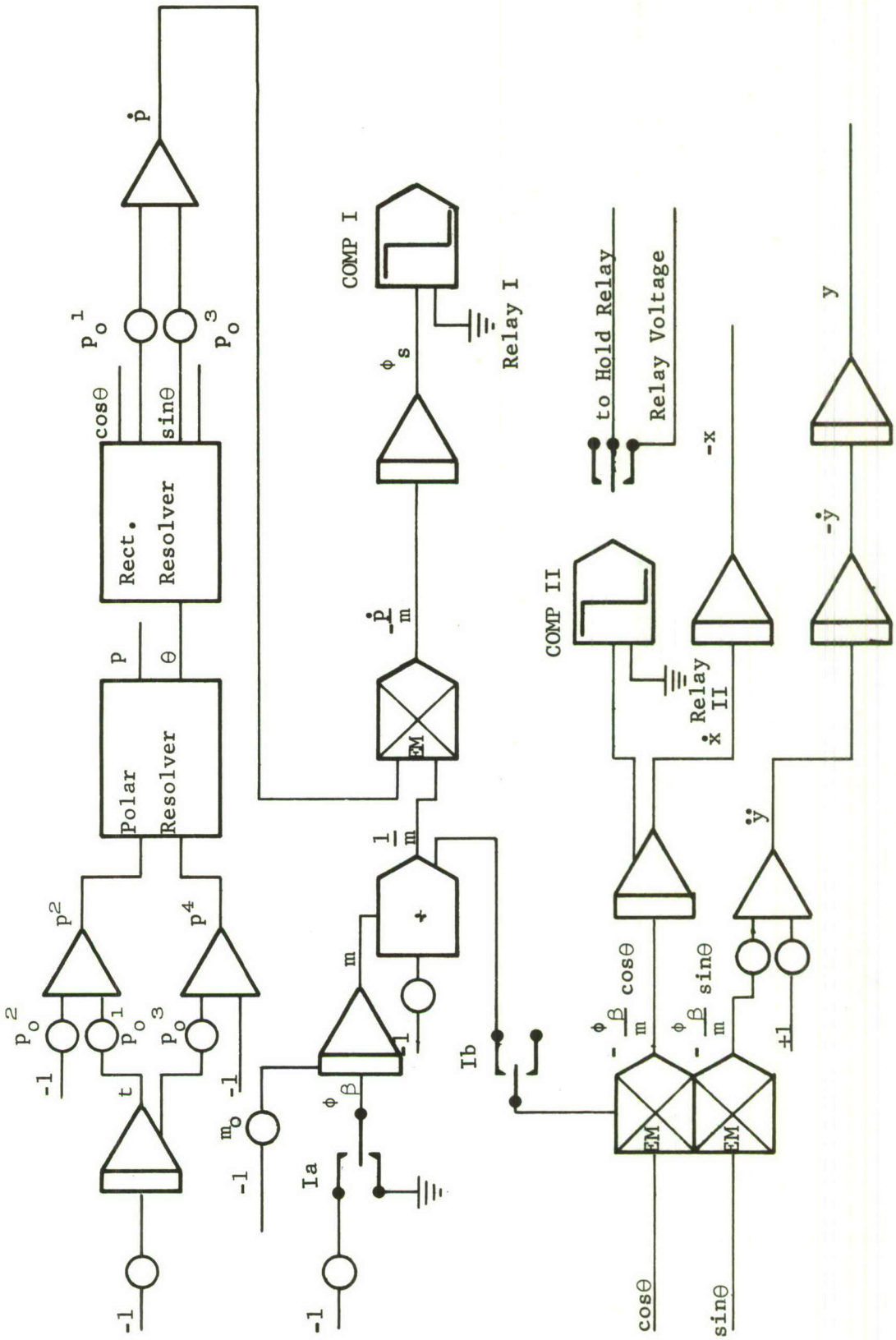
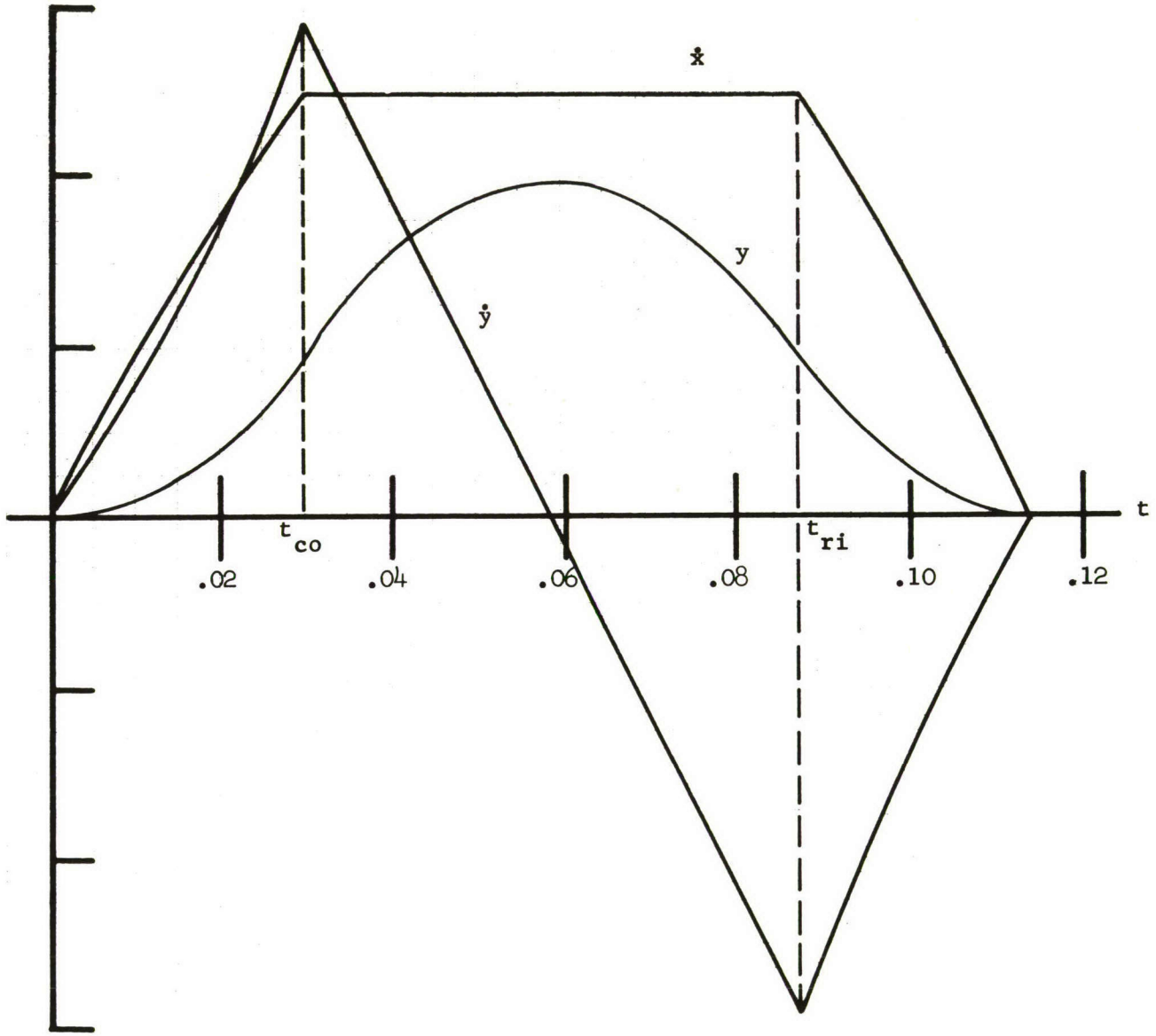


Fig. 2 Analog Computer Schematic



$$\begin{cases} \dot{x} = -.02 \\ \dot{y} = -.01 \\ y = -.0004 \end{cases}$$

$p_o^1 = 20$

$p_o^2 = 1.1496$

$p_o^3 = 0.3140$

Fig. 3 $\dot{x}$, $\dot{y}$, and y vs. time

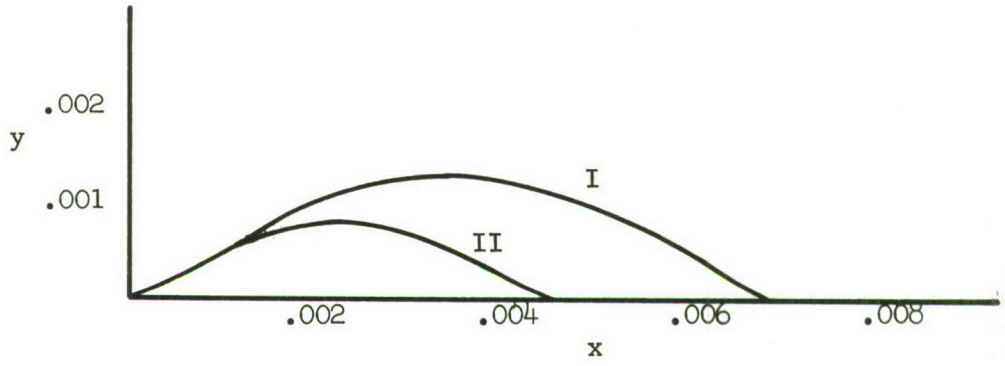
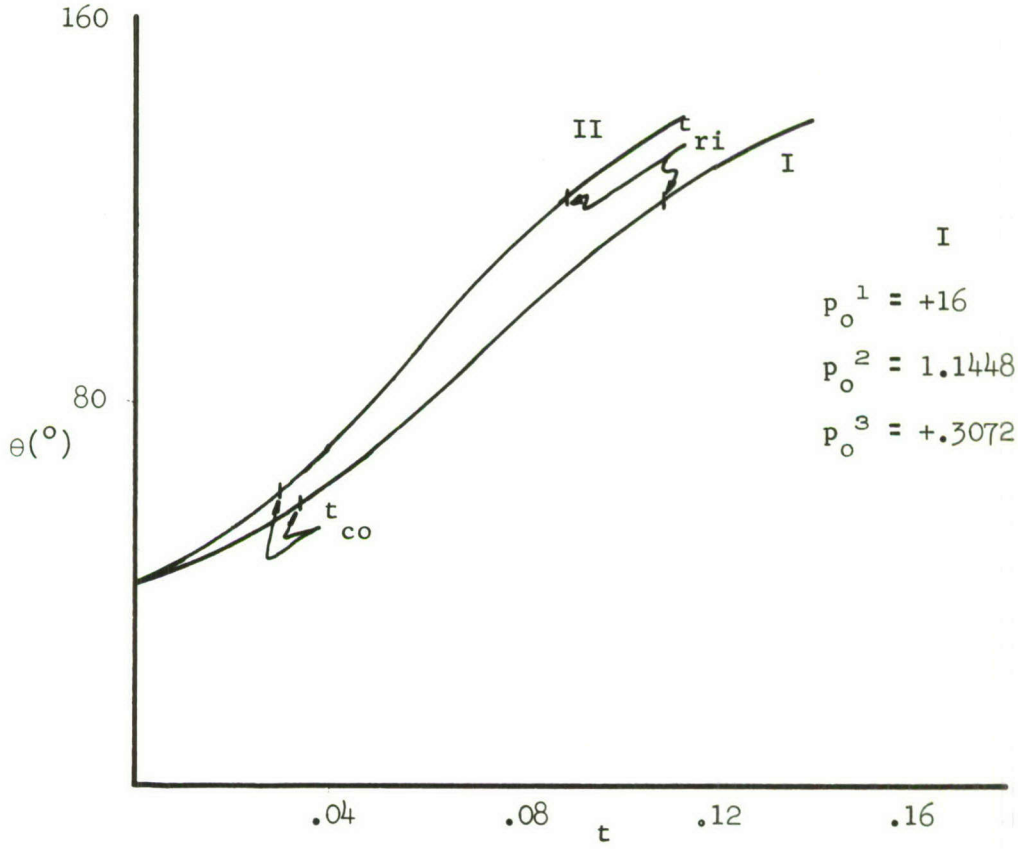


Fig. 4a Sample Trajectories $m_o = 0.4$



	I	II
p_o^1 =	+16	+20
p_o^2 =	1.1448	1.1498
p_o^3 =	+.3072	+.3140

Fig. 4b Thrust Angle vs. Time $m_o = 0.4$

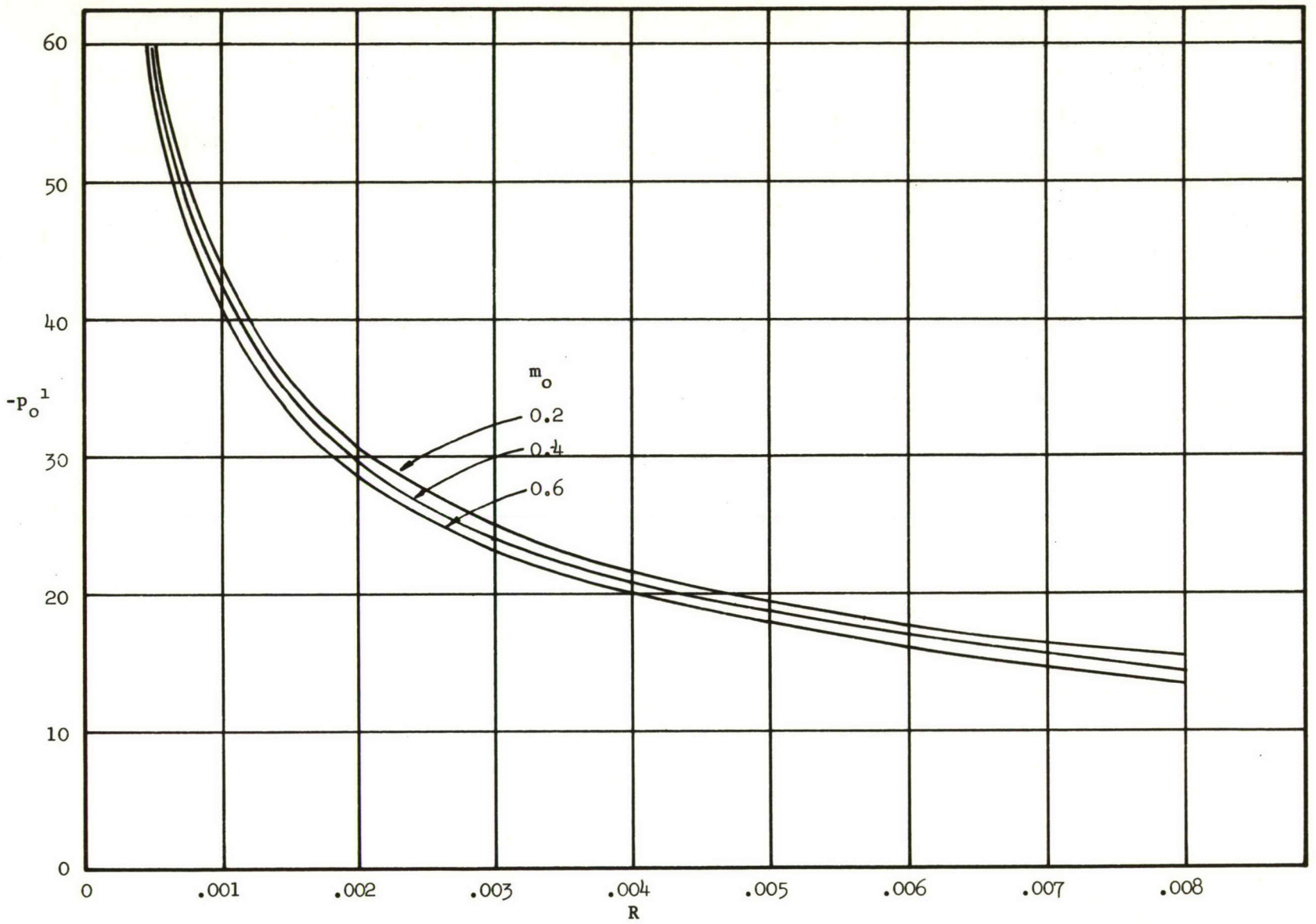


Fig. 5 $-p_o^1$ vs Range (R)

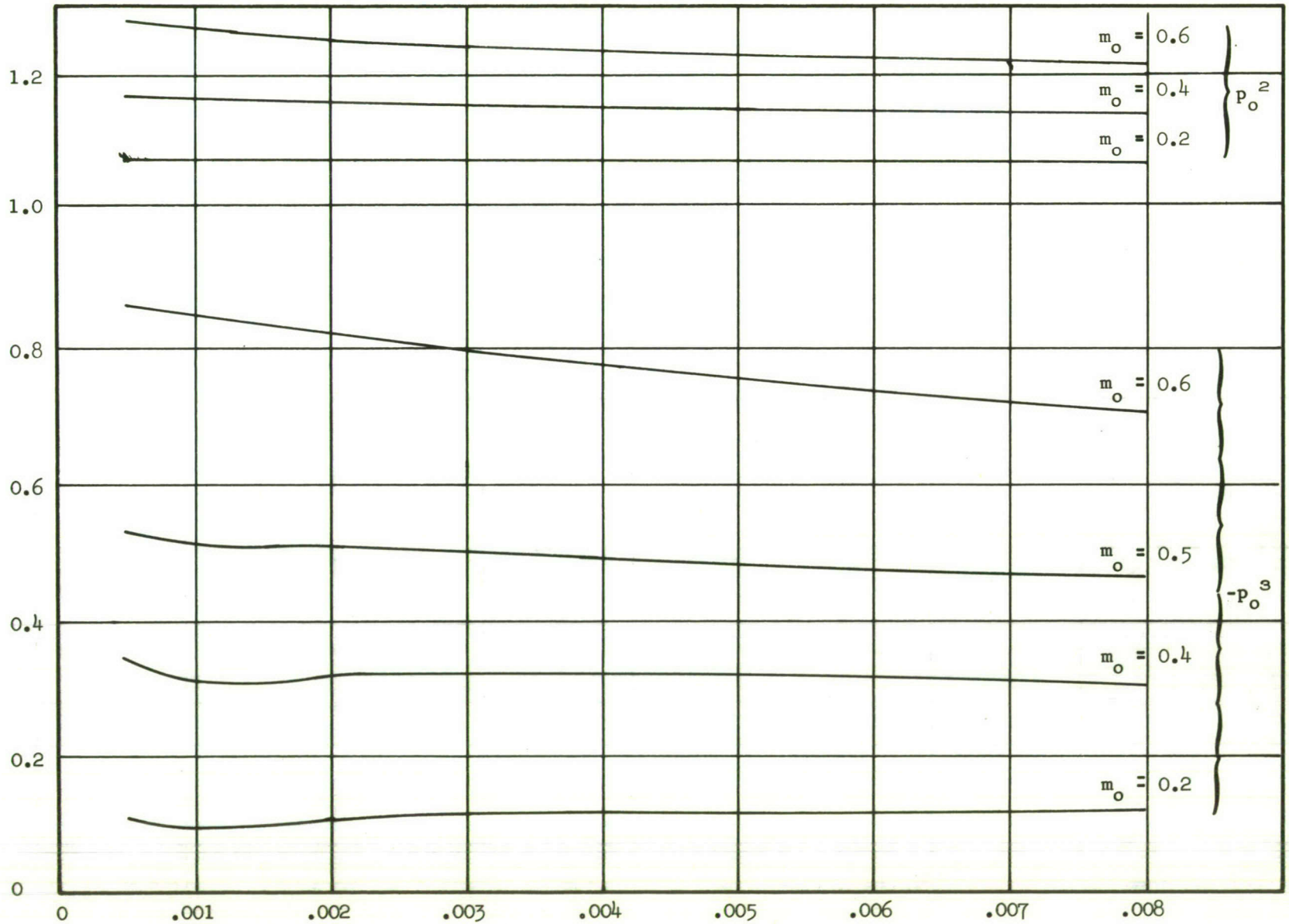


Fig. 6 p_o^2 and p_o^3 vs. Range (R)

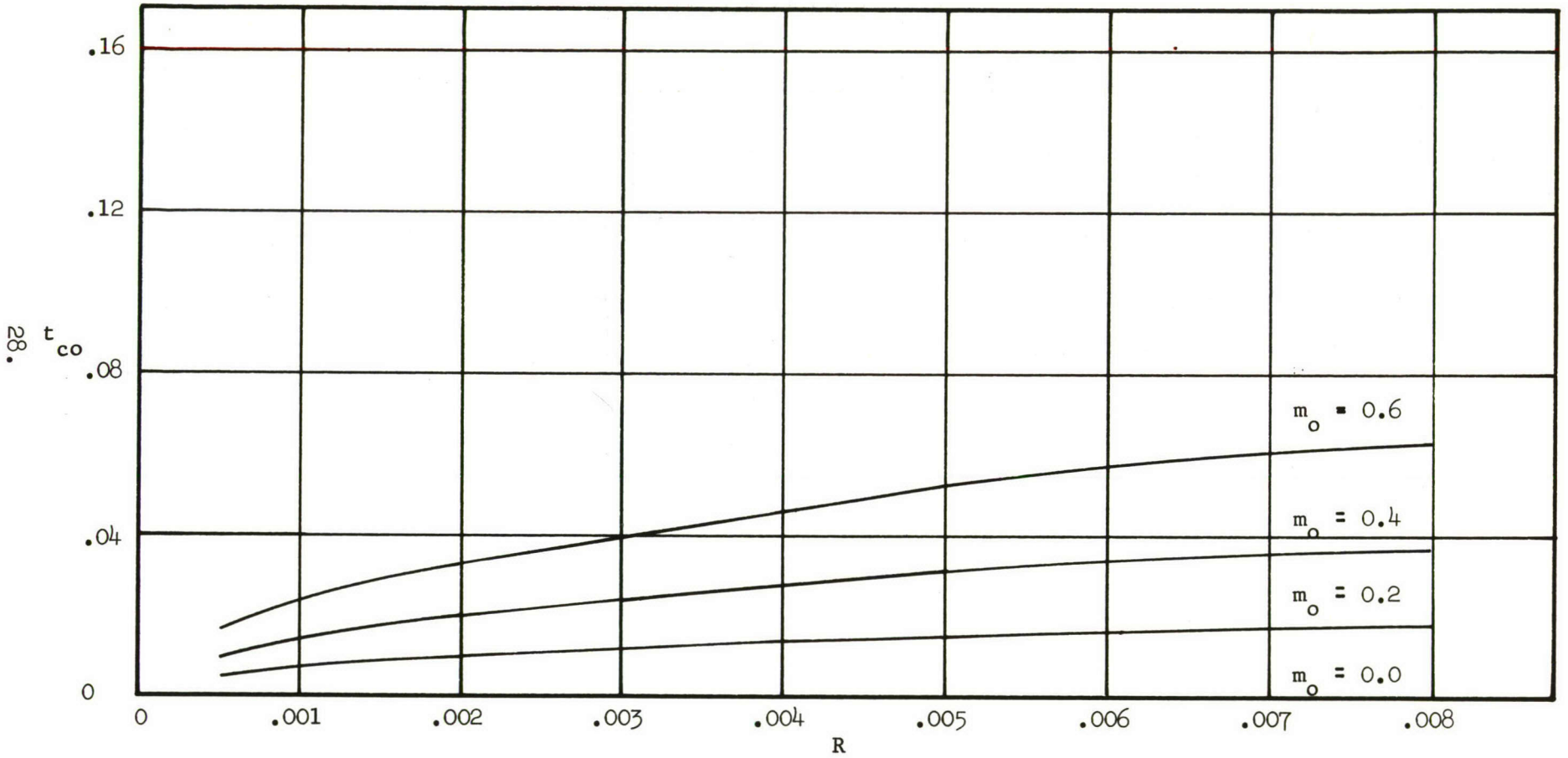


Fig. 7 Rocket Cut-off Time (t_{co}) vs. Range (R)

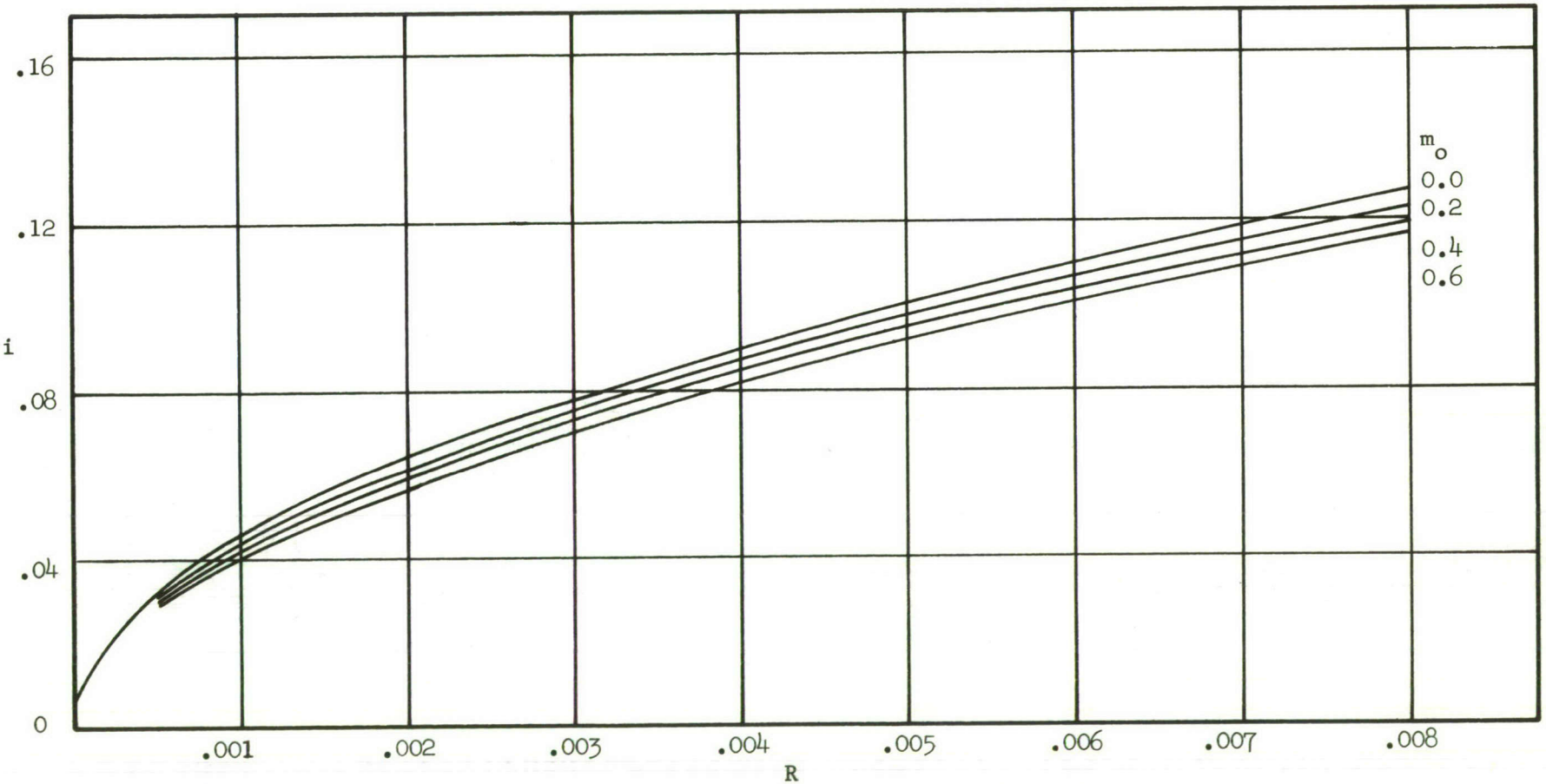


Fig. 8 Rocket Reignition Time (t_{ri}) vs. Range (R)

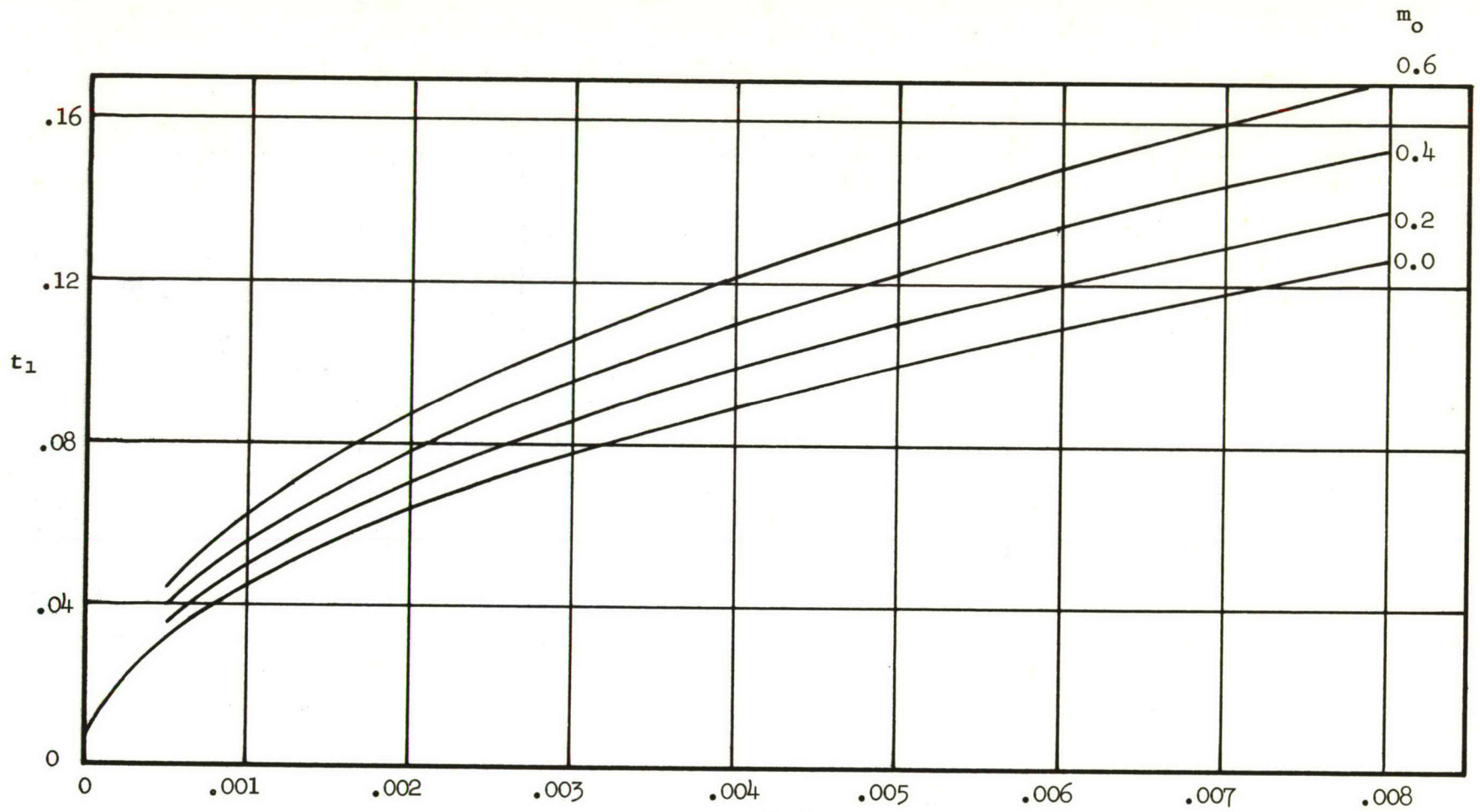


Fig. 9 Time of Landing (t_1) vs. Range (R)

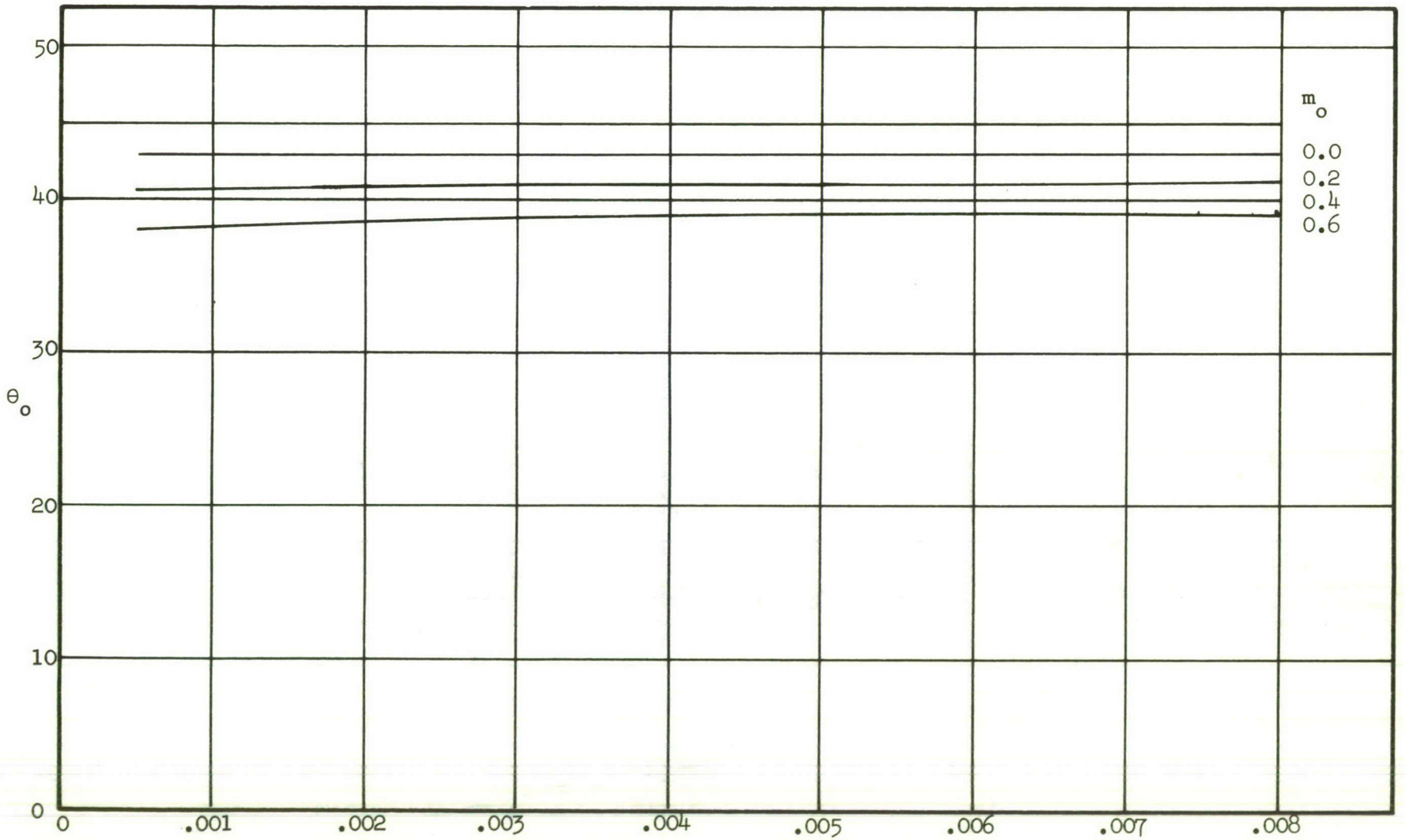


Fig. 10 Initial Thrust Angle (θ_o) vs. Range (R)

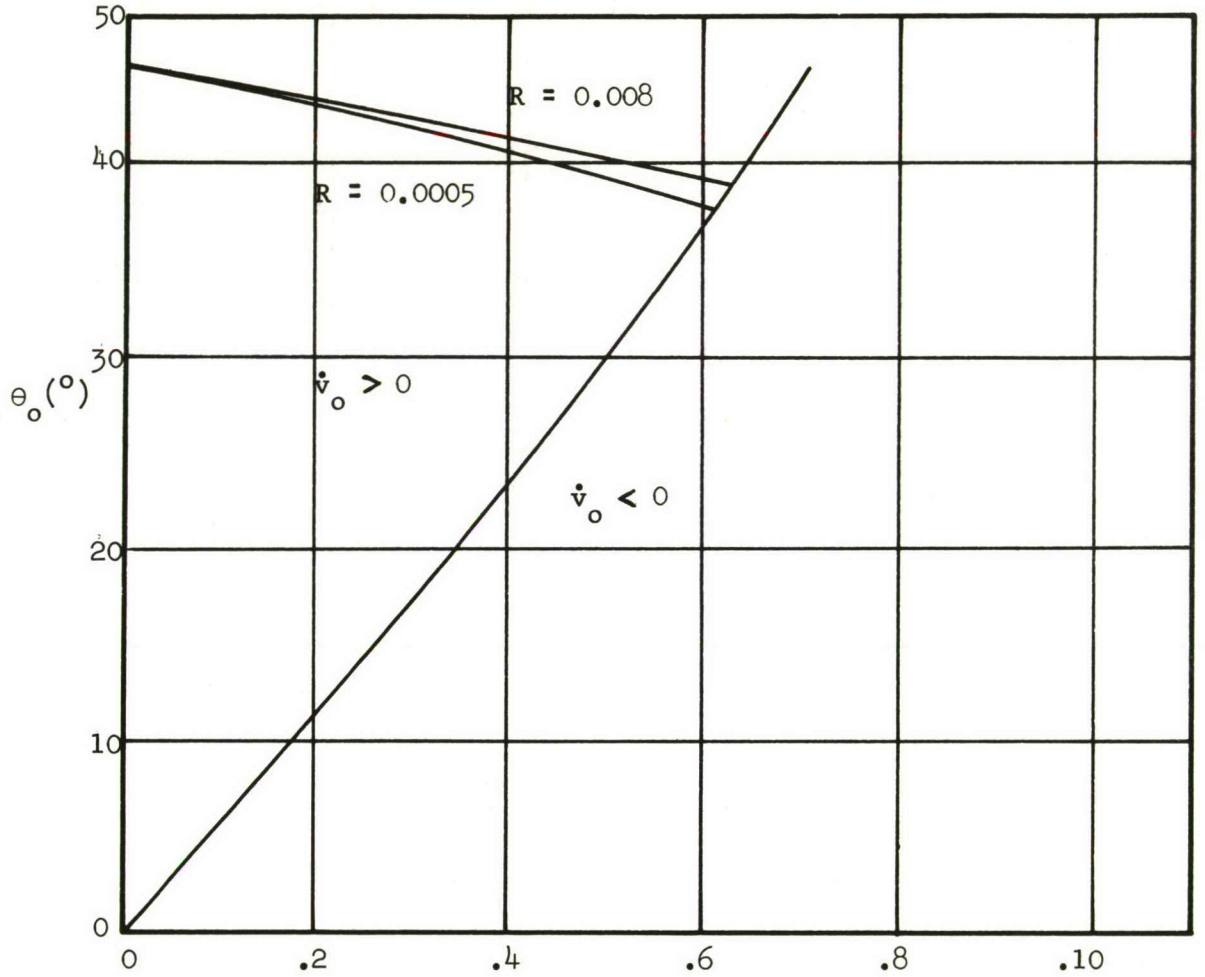


Fig. 11 Initial Thrust Angle (θ_o) vs. Initial Mass (m_o)

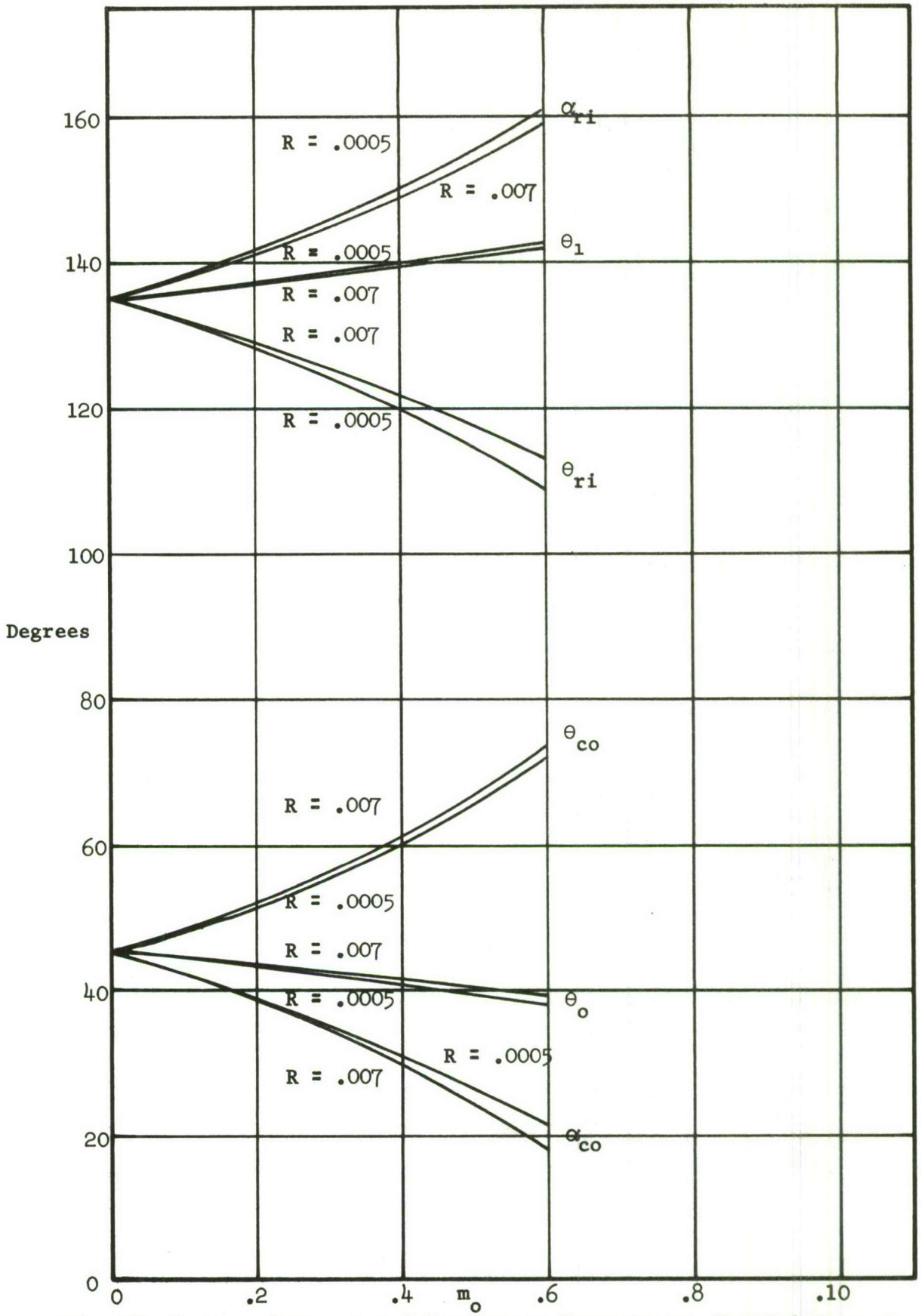


Fig. 12 Angles of Thrust and Trajectory Tangents vs. Initial Mass at $t = 0$, t_{co}, t_{ri}, and t_1

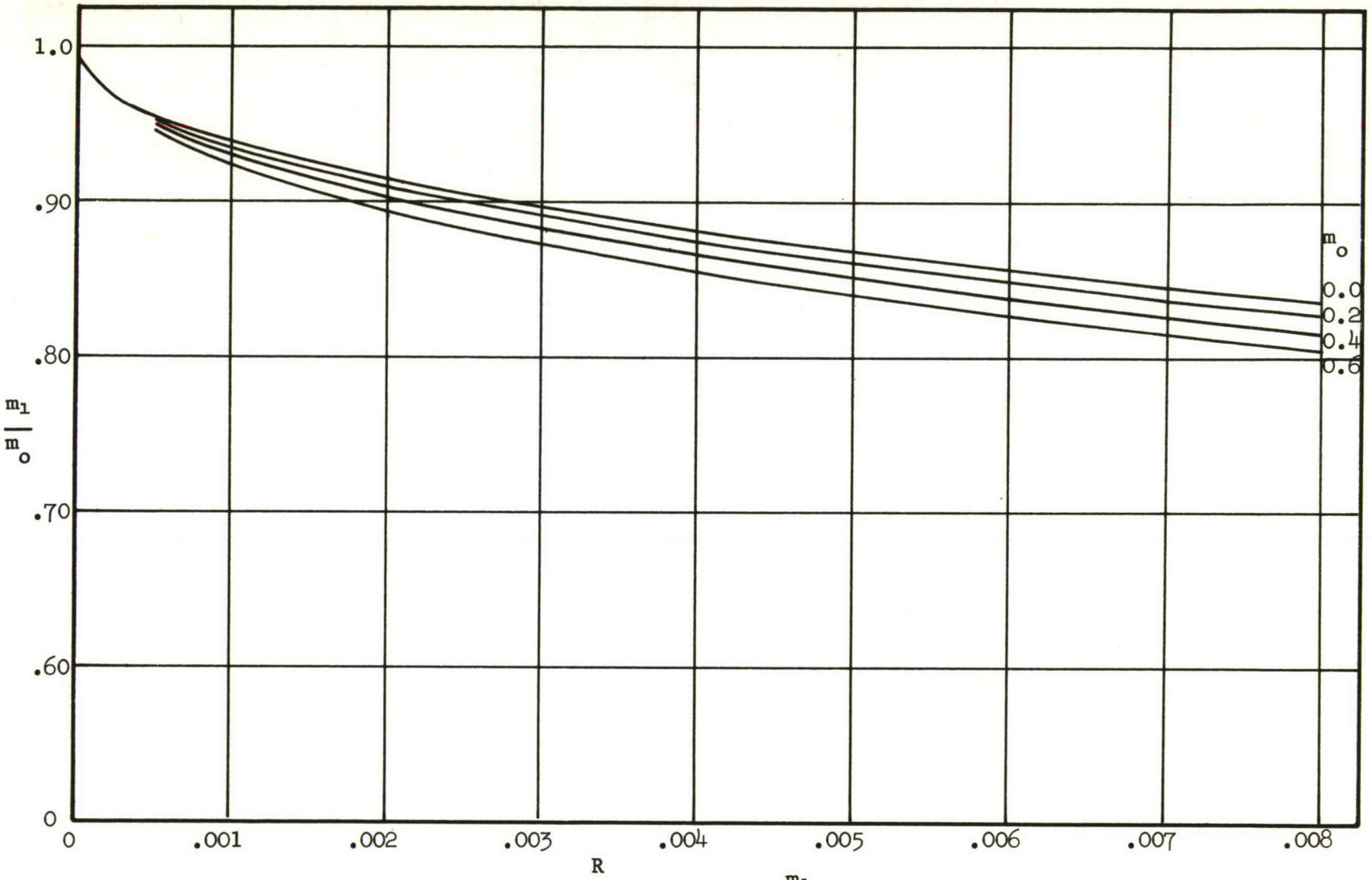


Fig. 13 Final to Initial Mass Ratio $\frac{m_1}{m_o}$ vs. Range

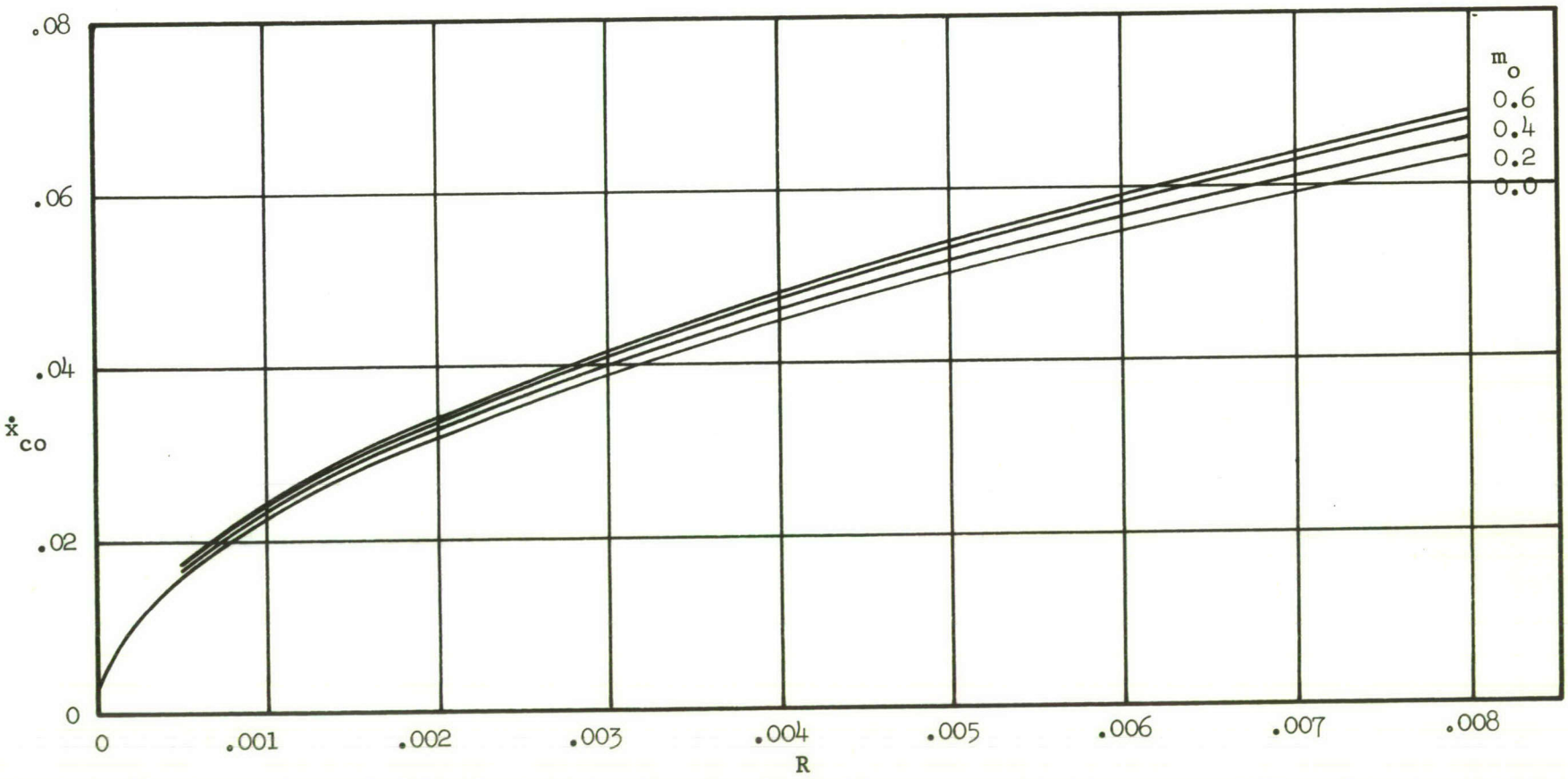


Fig. 14 Horizontal Velocity ($\dot{x}_{co}$) at Cutoff vs. Range (R)

References

1. Merrill, Grayson, ed., Principles of Guided Missile Design: Aerodynamics, Propulsion, Structures and Design Practice; D. Van Nostrand Company, Inc., Princeton, 1956.

2. Huntley, H.E., Dimensional Analysis, Rinehart and Co., New York, 1951.

3. Pontryagin, L.S.; Boltyanskii, V.G.; Gamkrelidze, R.V.; and Mishchenko, E.F., The Mathematical Theory of Optimal Processes, translated from the Russian by K. N. Trirogoff; Interscience Publishers, New York, 1962.

4. Leitmann, G., Optimization Techniques, Chapts 4 and 7, Academic Press, New York, 1962.

5. Rozonoer, L.I., "L.S. Pontryagin Maximum Principle in the Theory of Optimum Systems". Automat. i Telemekh. 20 (1959). (English translation, Automation Remote Control 1288-1302, 1405-1421, 1517-1532 (1960).

6. Valentine, F.A., in Contributions to the Calculus of Variations, 1933-1937; "The Problem of Lagrange with Differential Inequalities as Added Side Conditions," University of Chicago Press, Chicago, 1937.

7. Harnett, R.T., Sansom, F.J., and Warshawsky, L.M., Midas Programming Guide, Technical Document Report No. SEG-TDR-64-1, R&D, Wright-Patterson AFB, Ohio, January 1964.

Appendix I Solution for Impulsive Thrust

From Figure 1, the forces along the thrust vector may be written as

$$\frac{d\bar{v}_F}{d\bar{t}} = \frac{\bar{F}}{\bar{m}} - g \sin\theta$$

or

$$\frac{d\bar{v}_F}{d\bar{t}} = -\frac{c}{\bar{m}} \frac{d\bar{m}}{d\bar{t}} - g \sin\theta \qquad \text{(I-1)}$$

since

$$\bar{F} = c\beta = -c \frac{d\bar{m}}{d\bar{t}} .$$

If the mass starts from rest then the velocity along the thrust vector at cutoff is

$$\bar{v}_F(\bar{t}_{co}) = -c \int_{\bar{m}(\bar{t}_o)}^{\bar{m}(\bar{t}_{co})} \frac{d\bar{m}}{\bar{m}} - \int_{\bar{t}_o}^{\bar{t}_{co}} g \sin\theta \, d\bar{t} \qquad \text{(I-2)}$$

If the burning rate is allowed to increase without limit while the amount of fuel burned is held constant, the conditions for impulsive thrust are obtained. Thus

$$m(t_o) - \bar{m}(\bar{t}_{co}) = \beta(\bar{t}_{co} - \bar{t}_o) = \text{constant}$$

and

$$\lim_{\beta \to \infty} (\bar{t}_{co} - \bar{t}_o) = 0$$

Since the integrand of the second integral of (I-2) is bounded

$$\lim_{\beta\to\infty} \int_{\bar{t}_o}^{\bar{t}_{co}} g \sin\theta \, d\bar{t} = 0 .$$

while $$\lim_{\beta\to\infty} \int_{\bar{m}(\bar{t}_o)}^{\bar{m}(\bar{t}_{co})} \frac{d\bar{m}}{\bar{m}} = \ln \frac{\bar{m}(\bar{t}_{co})}{\bar{m}(\bar{t}_o)} .$$

The initial to final mass ratio for a velocity increment $\bar{v}_F(t_{co})$ which for impulsive thrust would be tangent to the trajectory is the familiar logarithmic function

$$\frac{\bar{v}_F}{c} = \ln \frac{\bar{m}(\bar{t}_o)}{\bar{m}(\bar{t}_{co})}$$

Introduction of the dimensionless ratios of equations (6) gives

$$v_F = \ln \frac{m(t_o)}{m(t_{co})} \qquad \text{(I-3)}$$

During the ballistic phase of the trajectory, the mass is a freely falling body so with an initial velocity v_F calculated from (I-3) and the equations of motion

$$\ddot{x} = 0$$

$$\ddot{y} = -1 .$$

The velocity at any time is

$$\dot{x} = v_F \cos\theta_o$$

$$\dot{y} = v_F \sin\theta_o - t$$

and the position is

$$x = (v_F \cos\theta_o)\, t$$

$$y = (v_F \sin\theta_o)\, t - \tfrac{1}{2}t^2 .$$

Since the gravity field is constant the trajectory is symmetrical with respect to the line $x = \frac{1}{2}(v_F \cos\theta_o)\ t_1$ where t_1 is the time at which body strikes the ground or

$$R = (v_F \cos\theta_o)\, t_1 .$$

At $\frac{1}{2}t_1$, the y velocity is zero and $y = y_{max}$. Thus

$$t_1 = 2v_F \sin\theta_o$$

and

$$R = 2v_F^2 \sin\theta_o \cos\theta_o \qquad \text{(I-4)}$$

The maximum range for a given v_F is obtained when $\frac{\partial R}{\partial \theta_o} = 0$. This relationship yields the familiar fact that $\theta_o = 45^o$ for maximum ballistic range in a vacuum.

For maximum range then the following expressions are obtained

$$R_{max} = v_F^2$$

$$y_{max} = \tfrac{1}{4} v_F^2 \qquad \text{(I-5)}$$

$$\frac{m(t_1)}{m(t_o)} = \exp\left(-2\sqrt{R_{max}}\right)$$

These equations were used to obtain the curves for $m_o = 0$ in Figures 7 through 14.

Acknowledgement

The author is indebted to Colonel R.R. Bate, Professor of Astronautics, and Captain R.D. McKelvey, Department of Astronautics, United States Air Force Academy, for suggesting this problem and for many helpful discussions, and to the Air Force Flight Test Center (AFSC) for making the analog computers of the Analog Simulation Division available for this study.

RONALD LEE LONG 1st Lt.

AFML (MANE) Wright-Patterson Air Force Base
OHIO

NEW SELECTIVELY ABSORBING COATINGS FOR SOLAR POWERED THERMOELECTRIC DEVICES (Unclassified)

TABLE OF CONTENTS

I. Introduction

II. System and Materials Considerations

A. System Aspects of the Problem

B. Coating Materials Aspects of the Problem

III. Experimental

A. Equipment

B. Coating Preparation

C. Results

IV. Discussion

V. Conclusions

VI. References

LIST OF ILLUSTRATIONS

1. Reflectance vs. wavelength of Al/FeOx $(3/4\ \lambda)/MgF_2$.
2. Reflectance vs. wavelength of Au/FeOx $(3/4\ \lambda)/MgF_2$.
3. Reflectance vs. wavelength for Al/FeOx $(1/4\ \lambda)$.
4. Reflectance vs. wavelength for Al/FeOx $(3/4\ \lambda)$.
5. Reflectance vs. wavelength for Au/FeOx $(1/4\ \lambda)$.
6. Reflectance vs. wavelength for Au/FeOx $(3/4\ \lambda)$.
7. Reflectance vs. wavelength for several single layer FeOx Coatings.
8. Reflectance vs. wavelength for Al/FeOx $(1/4\ \lambda)/MgF_2$ and Al/FeOx $(3/4\ \lambda)/MgF_2$.
9. Reflectance vs. wavelength for Al/FeOx $(3/4\ \lambda)/MgF_2$ and Au/FeOx $(3/4\ \lambda)/MgF_2$.
10. Reflectance vs. wavelength for Al/FeOx $(3/4\ \lambda)/MgF_2$ and Al/SiO/Al/SiO.
11. Solar absorptance (α) and total emittance (ε) vs. temperature for Al/FeOx $(3/4\ \lambda)/MgF_2$.
12. Reflectance vs. wavelength for Al/FeOx $(3/4\lambda)/MgF_2$ before and after 100 hours ultraviolet exposure at a pressure of 2×10^{-5} Torr.

I. Introduction

Space vehicle power requirements have created a need for developing devices which will increase the electrical output of present auxiliary power systems. In addition, other desirable improvements include resistance to radiation damage, good thermal cycling properties and longer life.

Some power systems, specifically, solar powered heat engines, have better radiation resistance, operate at higher power levels, and potentially have a longer life than presently available solar cells. However, the efficiencies of these devices are much lower. Many papers discussing the operating principles and advantages and disadvantages of these devices appear in the literature, hence, these topics will not be discussed here (ref. 1-4).

This paper deals with increasing the electrical output of thermoelectric auxiliary power systems by directly increasing the efficiency of the collector surface. The collector surface is made of a thin metal sheet, usually aluminum. By applying a selectively absorbing coating to the collector surface, the heat collection efficiency (consequently, the power output) can be increased considerably.

II. System and Materials Considerations

A. System Aspects of the Problem

The efficiences of thermoelectric devices, as shown by Cross (ref. 5) and Schmitt (ref. 6), can be increased by applying a selective absorber to the collector surface. A selective absorber is a coating which absorbs and retains as much incident solar energy as possible, i.e., it possess a high solar absorptance (α) and low infrared emittance (ε). The coating increases the energy absorbed while the low emittance metal of which the collector is made keeps the amount of reradiated energy to a minumum. The theory of selective absorbers is dealt with very well by the preceeding two references, and will not be discussed in detail.

The overall efficiency of these devices can be considered as a product of the collector efficiency and the Carnot or heat engine efficiency. The Carnot efficiency is a function of the collector and radiator temperatures, and collector efficiency is a function of the solar absorptance and the total emittance (at a given operating temperature). The following equation expresses the fundamental relationship between the power output (W) and the absorber and Carnot efficiencies.

$$W = (\alpha n G - \sigma \epsilon T_1^4)(1 - T_2/T_1)$$

where

W = power output/cm^2 of collector surface

α = solar absorptance

n = concentration factor (equal to 1 if no concentrating mirror is used)

G = solar constant above the atmosphere

ϵ = total emittance of the collector surface

σ = Stefan-Boltzmann constant

T_1 = collector operating temperature

T_2 = radiator operating temperature

From this equation we see that, given T_1, T_2 and n, the absorber efficiency is dependent on the absolute values of the solar absorptance and the total emittance.

B. Coating Materials Aspects of the Problem

As stated previously, coatings which possess a high solar absorptance and low infrared emittance are called selective absorbers and there exist a number of techniques for producing coatings with these properties (ref. 7). For use in space, the coating material itself should possess the following properties:

1. stability to a vacuum environment at a temperature well above the planned operating temperature; 2) stability to ultraviolet radiation; and 3) a low vapor pressure to prevent evaporation in a vacuum.

Iron oxide (FeOx) is a material which possess a very low vapor pressure and has a relatively high melting point. In bulk form it possesses good absorption characteristics. It was hoped by combining these properties to take advantage of optical"interference" effects that a superior solar absorber might be found. Since prototype thermoelectric convertors use an aluminum collector, this material was selected as the substrate material. This laboratory has also found that fired-on gold possesses a very low emittance and is easy to apply. Therefore, gold coated aluminum was also considered as a substrate material.

III. Experimental

A. Equipment

The equipment used in this research consisted of vacuum evaporator, a vacuum ultraviolet radiation chamber and a spectroreflectometer. The vacuum evaporator was equipped with a 1400 L/sec oil diffusion pump backed by a 30 cfm mechanical pump and was used for preparing all coatings. To minimize backstreaming (necessary for good optical films) the diffusion pump was trapped with an optically dense dual coolant baffle and a cooling cap. In addition, an automatic liquid nitrogen transfer control device was constructed and employed. This maintained the trap at the boiling point of liquid nitrogen (77°K).

A pyrex bell jar housed the necessary electrical and mechanical feed-thrus, optical system, glow discharge cleaning rings, work holder and substrate heater. A 5000v, 300 ma power supply was used for glow discharge cleaning.

A Welch Densichron optical system was used to monitor the substrate reflectance during the evaporation. This instrument consisted of a projector with a tungsten bulb, a photocell sensitive to the projector output and a voltage stabilized response meter. During the evaporations, the substrate reflectance passed through a series of maxima and minima. These were indicated on the response meter and the evaporations halted with a shield when the proper thickness was achieved.

The vacuum ultraviolet test chamber, the spectroreflectometer, and the procedures used with these instruments have been described previously by Hormann (ref 8).

B. Coating Preparation

Using aluminum and gold (fired-on aluminum) as substrates, several films of iron oxide were vacuum deposited. Although the well known one-quarter wavelength for minimum reflectance does not hold precisely for films with an absorbing substrate (ref 9), first and second reflection minima of iron oxide are indicated by $(1/4\lambda)$ and $(3/4\lambda)$, respectively. The evaporated iron oxide films consisted of an unknown suboxide of iron and are indicated by FeOx. Tungsten filaments for evaporating the iron oxide were prepared by dipping them into a paste of Fe_3O_4 (C.P.) and drying by resistance heating.

Magnesium fluoride was evaporated from a conventional MgF_2 boat. Whenever it was used, magnesium fluoride was evaporated until the first minimum in reflectance occurred. In all cases, this minimum was lower than that for the single layer FeOx films.

The substrates were cleaned by washing in detergent, rinsing in distilled water and then in methanol. They were then placed in the vacuum chamber and cleaned by ion bombardment for approximately ten minutes at a pressure of about 50 microns. The chamber was then evacuated to 2×10^{-6} Torr and the filaments outgassed. During evaporation of iron oxide, the pressure rose to the 10^{-5} Torr range due to some decomposition of the material at its evaporation temperature. The pressure was maintained on the 10^{-6} Torr range for all magnesium fluoride evaporations.

C. Results

The nature of the Welch Densichron required determination of its reproducibility when used as a film thickness monitor. Figures1 and 2 show the reflectance of several different samples of each of the two best coatings. Considering the equipment used,

reasonable reproducibility was obtained.

Figures 3-6 show the reflectance (0.4 - 0.7 microns) of several single layer coatings of Al/FeOx and Au/FeOx, while Figure 7 shows a comparison of the best of each. One can readily observe that the thin (1/4 λ) iron oxide films are superior to the thicker (3/4 λ) films. Some single layer coatings were found to oxidize on long term exposure (about 1 month) to ambient room temperature conditions, although other specimens seem indefinitely stable. To inhibit oxidation an overcoat of MgF_2 was evaporated over the iron oxide film. Figure 8 shows the reflectance (0.4 -0.7 microns) of Al/FeOx (1/4 λ)/MgF_2 and Al/FeOx (3/4 λ)/MgF_2 multilayer films. In this case, the coating with the thickest iron oxide film gives the best results. Figure 9 shows a comparison of the same coating on aluminum and gold substrates. From this figure the Al/FeOx (3/4 λ)/MgF_2 coating is readily observed to be superior to the Au/FeOx (3/4 λ)/MgF_2 coating. Figure 10 shows the reflectance of the Al/FeOx (3/4 λ)/MgF_2 coating from 0.4 - 15 microns. A dotted line indicates and extrapolated region where no actual measurements were available. For comparison purposes, the reflectance of the coating being used on prototype converters is included in the same figure (ref. 10). The measured solar absorptance and total emittance of the Al/FeOx (3/4 λ)/MgF_2 coating is shown in Figure 11.

The vacuum ultraviolet stability of the Al/FeOx (3/4 λ)/MgF_2 coating was determined by irradiating the coating with simulated concentrated solar energy for 100 hours at a pressure of 2×10^{-5} Torr. During the exposure, the sample reached a temperature of 150°C. Figure 12 shows the reflectance (0.4 - 0.7 microns) of this coating before and after exposure. Incomplete vacuum-thermal studies indicate that when a good overcoat of MgF_2 is obtained and the sample immediately placed in the vacuum furnace the coating is stable to approximately 500°C.

IV. Discussion

Reflectance data is often used for initial evaluation of coating materials rather than the time consuming and tedious emittance and absorptance measurements. In this investigation it has been found that solar absorptance and total emittance values calculated from reflectance data are good approximations of the measured values. Therefore absorptance and emittance measurements were obtained only for the coating with the best reflectance.

Two unexpected results were obtained during this research. First, since the thin ($1/4\ \lambda$) single layer coatings were superior to the thick ($3/4\ \lambda$) coatings, it was anticipated that Al/FeOx ($1/4\ \lambda$)/MgF_2 would be superior to Al/FeOx ($3/4\ \lambda$)/MgF_2. Such was not the case and an explanation for this is not available. However, iron oxide is a selective absorber in itself (called a selective black). And, in all the evaporations no attempt was made to obtain a particular thickness but emphasis was placed on obtaining the lowest reflectance possible. In addition, the single layer theory, for transparent media cannot be expected to hold for multilayer films on absorbing media.

Secondly, the reflectance of both single layer and multilayer coatings was improved by exposure to ultraviolet radiation. Figure 12 shows a decrease in the visible reflectance after 600 equivalent sun hours and it is expected that the infrared reflectance will be increased thereby producing overall improvement in the original sample. As far as the author knows, this is the first time this phenomenon has been reported. This, then, suggests a novel method for preparing of improving selective absorbers by exposing coatings (which might be rejected) to intense ultraviolet radiation.

V. Conclusions

The major accomplishment of this research was the preparation of a new coating which extends the temperature range of presently available solar absorbers for

thermoelectric devices. This coating has the potential of giving comparable results as the presently used coating yet at a much higher temperature. The major disadvantage of this coating is its infrared emittance. However, further research using different techniques and better thickness control should reduce the emittance considerably. Specific approaches to accomplish this end are as follows:

1. Use of a semitransparent metal spacer for this coating.
2. Use of a more accurate thickness monitor (such as a quartz oscillator) to improve reproducibility.
3. Investigation of the influence of other topcoat materials (such as CeO_2) on the reflectance of the system.
4. Investigation of the influence of ultraviolet radiation on the infrared reflectance of the system.

VI. References

1. Davisson, J. W., and Pasternak, J., "Status Report on Thermoelectricity." NRL Memorandum Report No. 1361., U.S. Naval Research Laboratory, Washington, D.C., October 1962.

2. Hemle, A., et al., "Flat Plate Thermoelectric Panels" ASD-TDR-62-214. Aeronautical Systems Division, Wright-Patterson Air Force Base, Ohio, May 1962.

3. "Advanced Solar Thermoelectric Power Systems", ASD-TDR-62-877. Aeronautical Systems Division, Wright-Patterson Air Force Base, Ohio, December 1962. (Prepared by Thompson Ramo Woolridge, Inc., Cleveland, Ohio).

4. Beggs, J. E., et al., "Research on Thermionic Converters" AFCRL 430. Air Force Cambridge Research Laboratory, Bedford, Massachusetts.

5. Cross, R.A., "Properties Required for Collector Surfaces of Solar-Powered Thermal Systems", ASD-TDR-62-919. Aeronautical Systems Division, Wright-Patterson Air Force Base, Ohio, April 1963.

6. Schmidt, R. N., " Effectiveness of Solar Absorber Surfaces", Presented at the A.R.S. Space Power Systems Conference, Santa Monica, California, September 1962.

7. Long, R. L., and Weaver, J. H., "Selective Absorbers for Solar Energy Collectors." Technical Memorandum ASRCN 63-4, Directorate of Materials and Processes, Wright-Patterson Air Force Base, Ohio, January 1963.

8. Hormann, H.H., "Thermal Control Coatings in a Space Environment", (to be published).

9. Schmidt, R. N., and Janssen, J.,"High Temperature Solar Absorber Coatings", RTD-TDR-63-579, Part I, Wright-Patterson Air Force Base, Ohio, November 1963.

10. Long, R. L., "Evaluation of the Hass Dark Mirror as a Selective Absorber". (to be published).

Fig. 1 REFLECTANCE vs. WAVELENGTH FOR $Al/FeOx(\frac{3}{4}\lambda)/MgF_2$

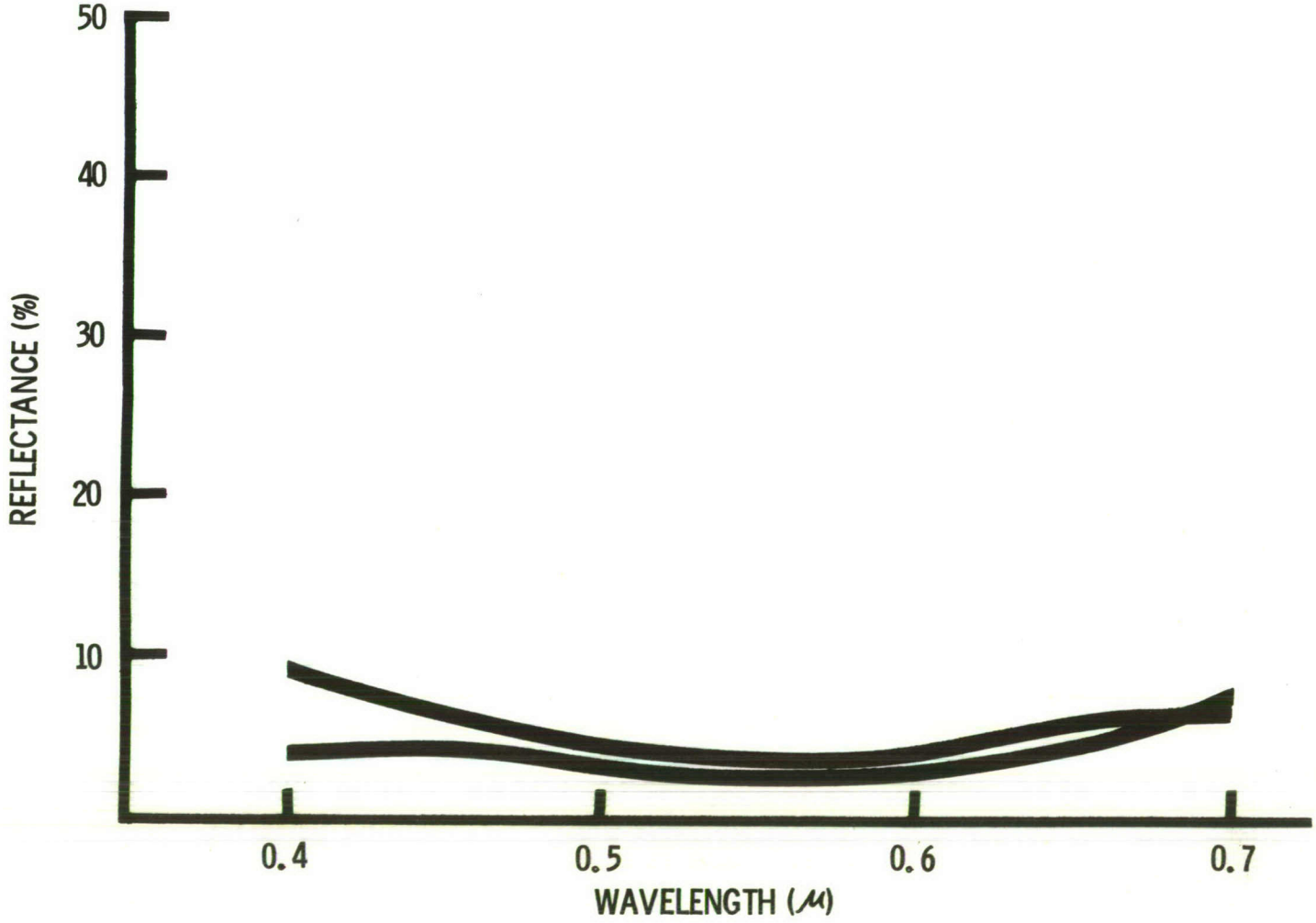


Fig. 2 REFLECTANCE vs. WAVELENGTH FOR $Au/FeOx(\frac{3}{4}\lambda)/MgF_2$

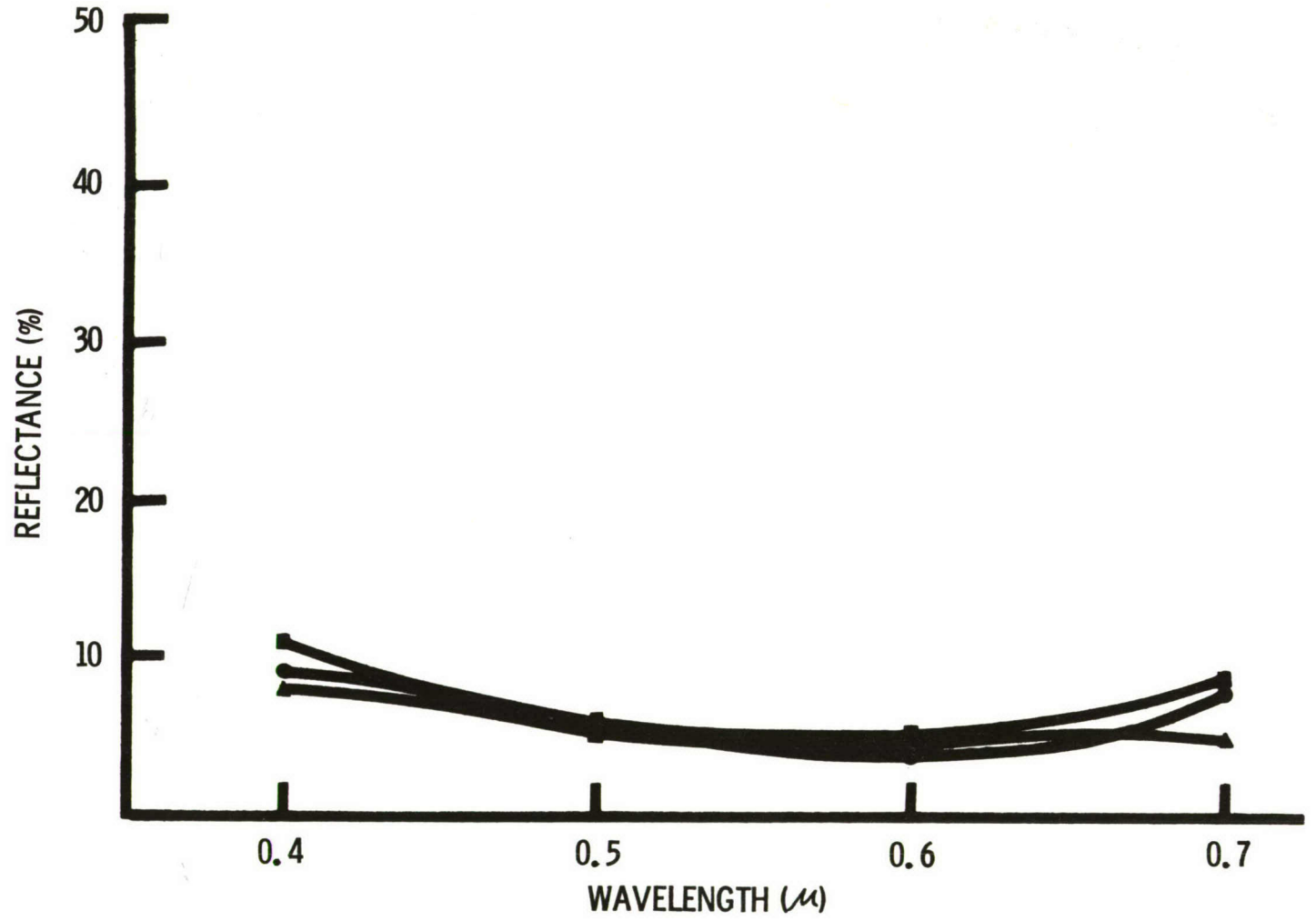


Fig. 3 REFLECTANCE vs. WAVELENGTH FOR Al/FeOx(¼λ)

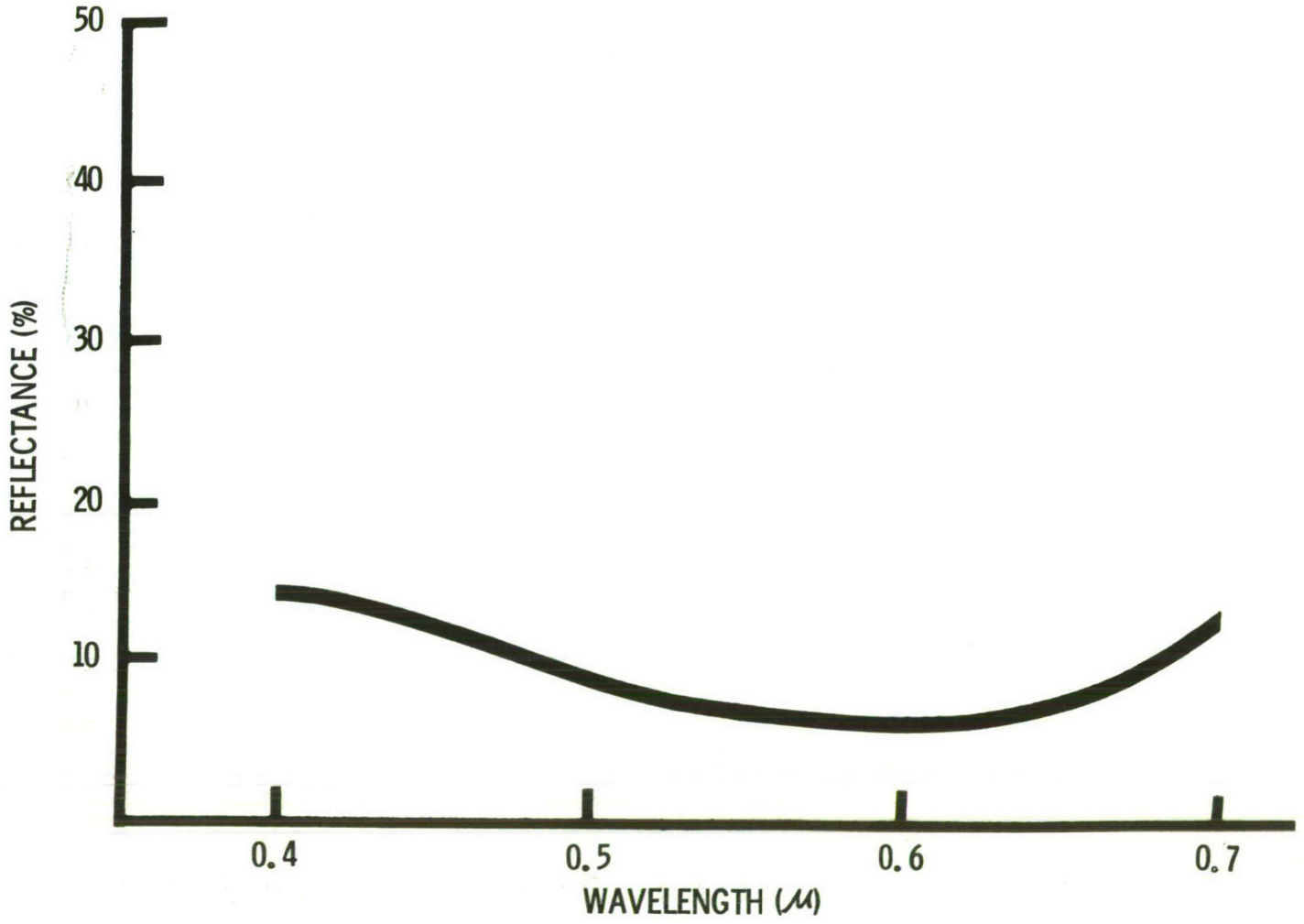


Fig. 4 REFLECTANCE vs. WAVELENGTH FOR Al/FeOx(¾λ)

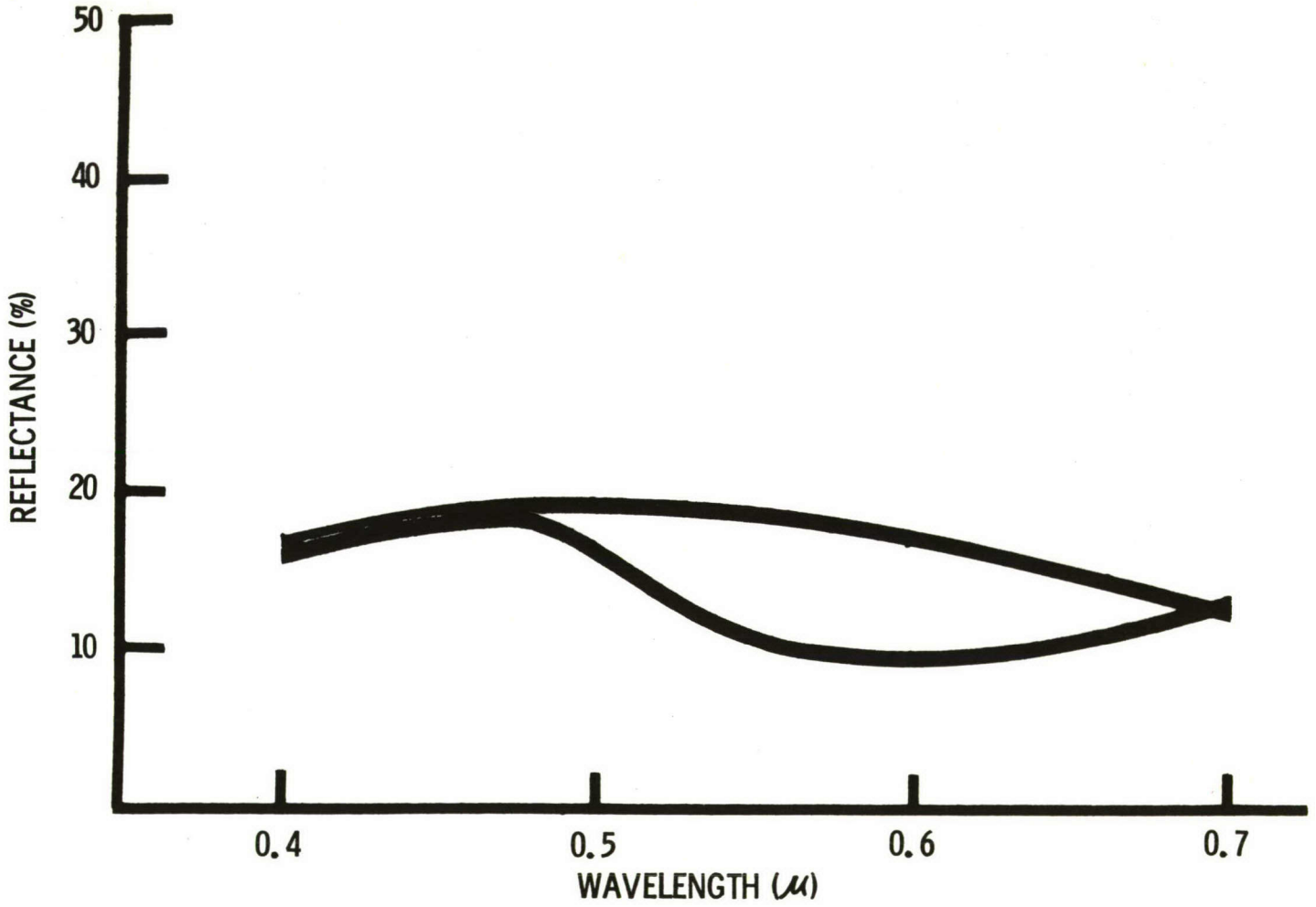


Fig. 5 REFLECTANCE vs. WAVELENGTH FOR Au/FeOx(¼λ)

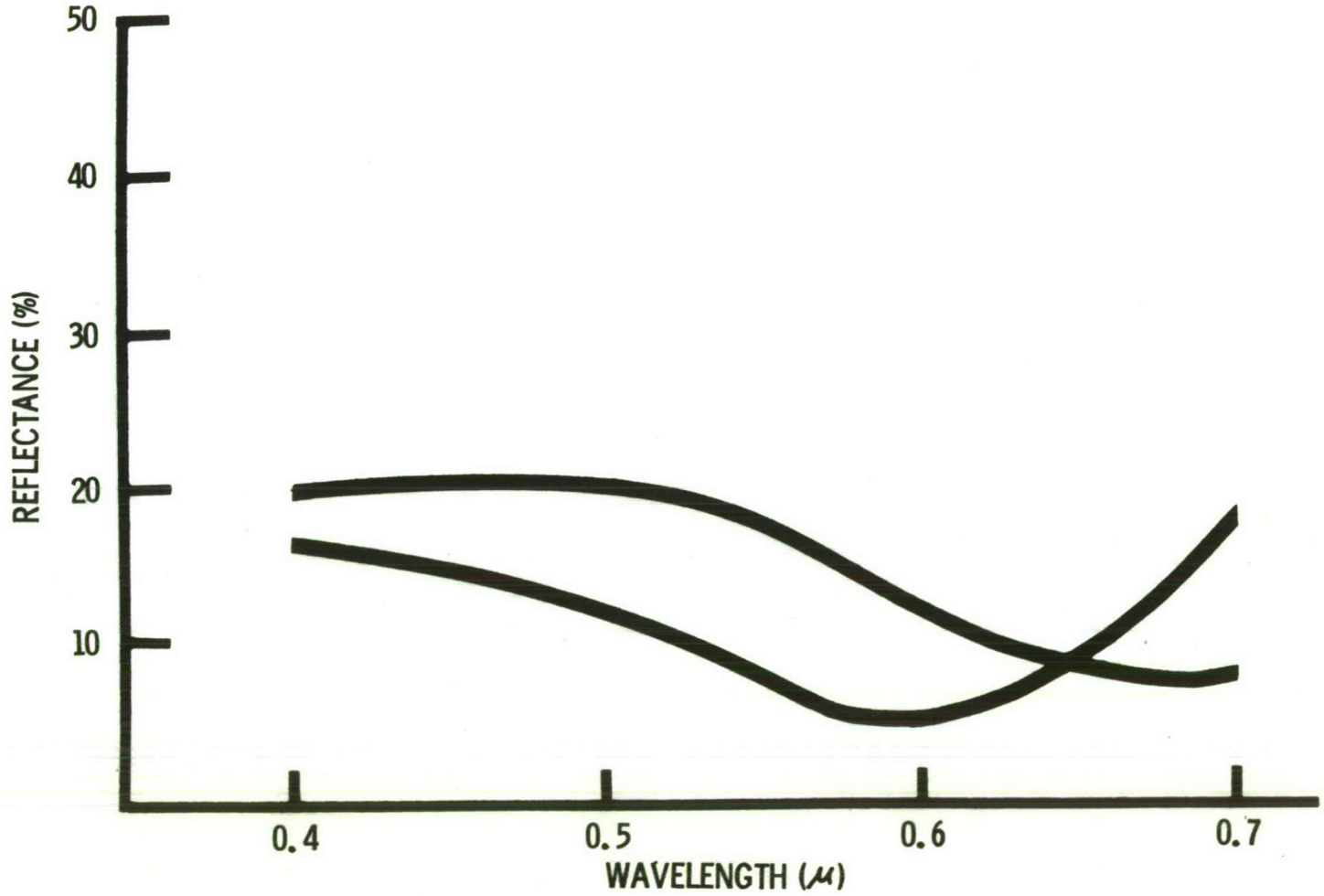


Fig. 6 REFLECTANCE vs. WAVELENGTH FOR Au/FeOx(¾λ)

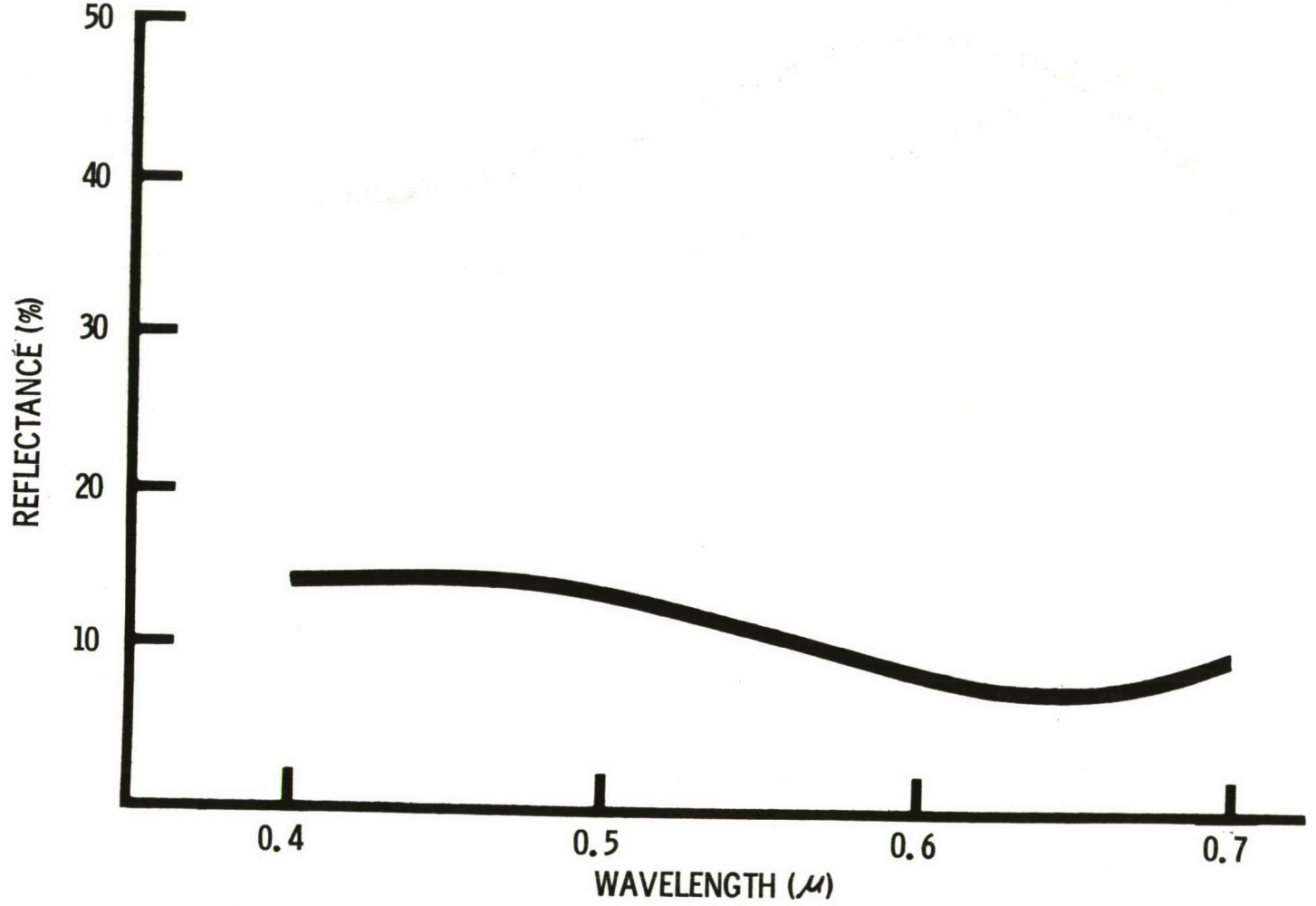


Fig. 7 REFLECTANCE vs. WAVELENGTH OF SEVERAL SINGLE LAYER FeOx COATINGS

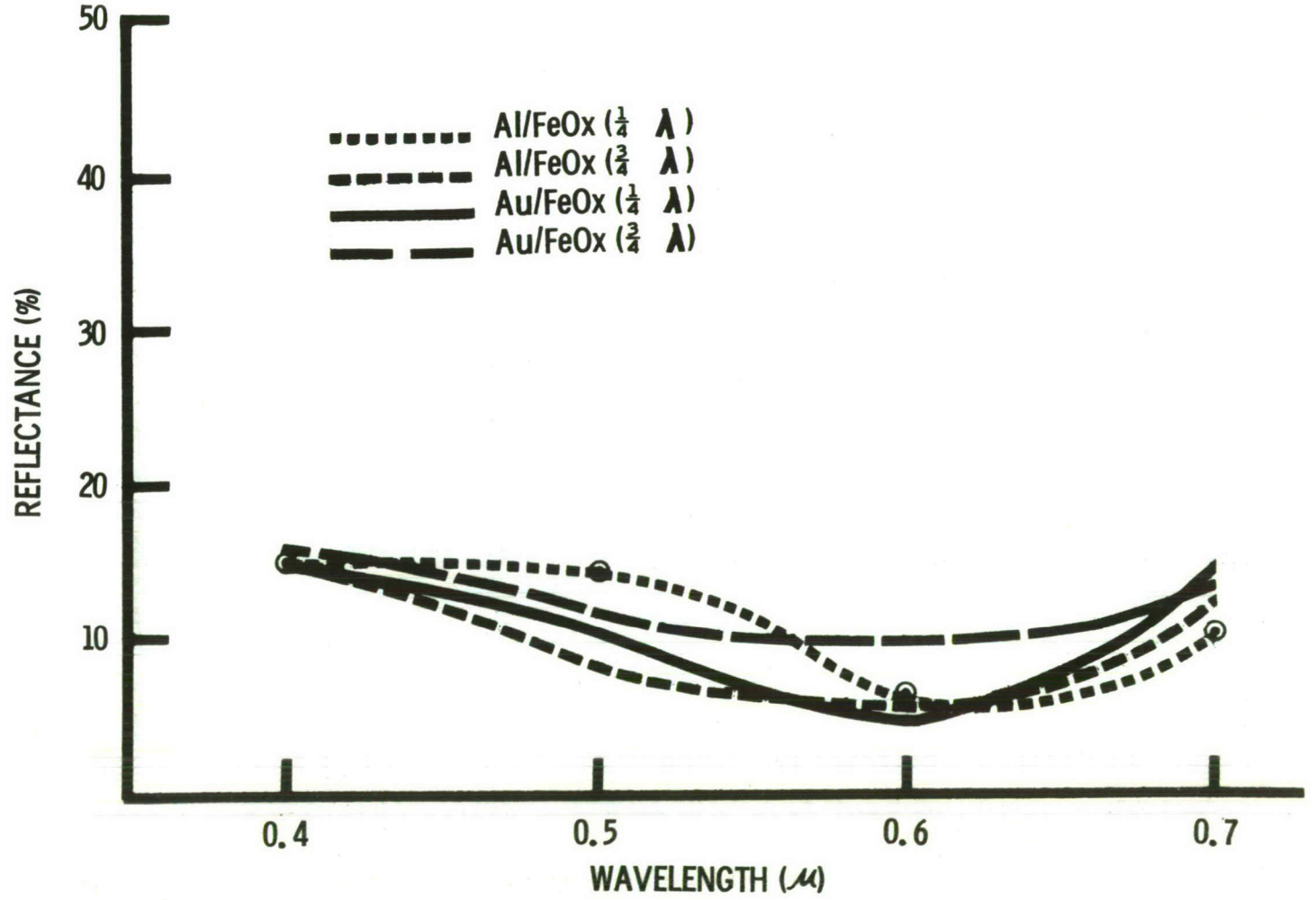


Fig. 8 REFLECTANCE vs. WAVELENGTH FOR Al/FeOx(¼λ)/MgF_2 AND Al/FeOx(¾λ)/MgF_2

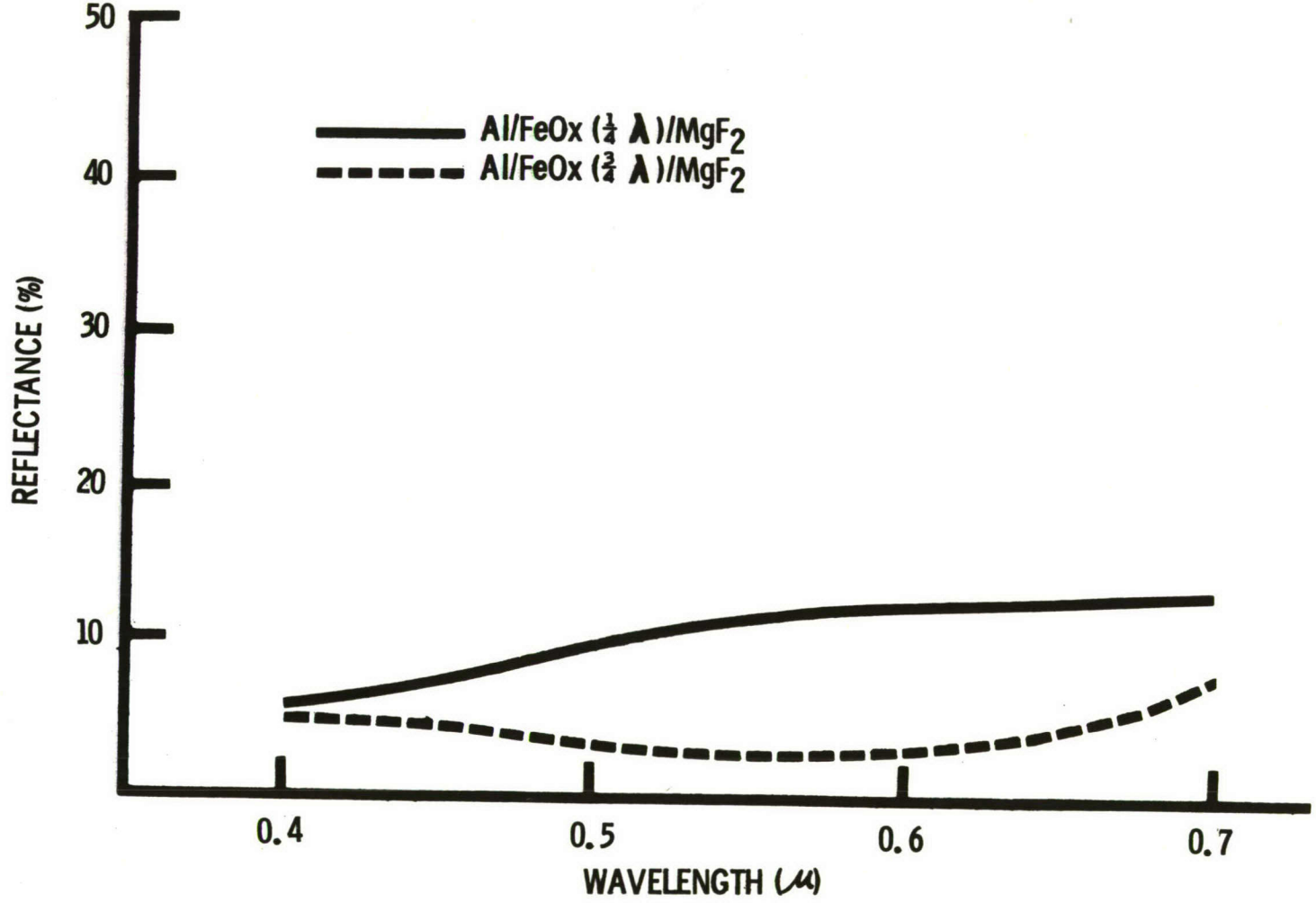


Fig. 9 REFLECTANCE vs. WAVELENGTH FOR Al/FeOx(¾λ)/MgF_2 AND Au/FeOx(¾λ)/MgF_2

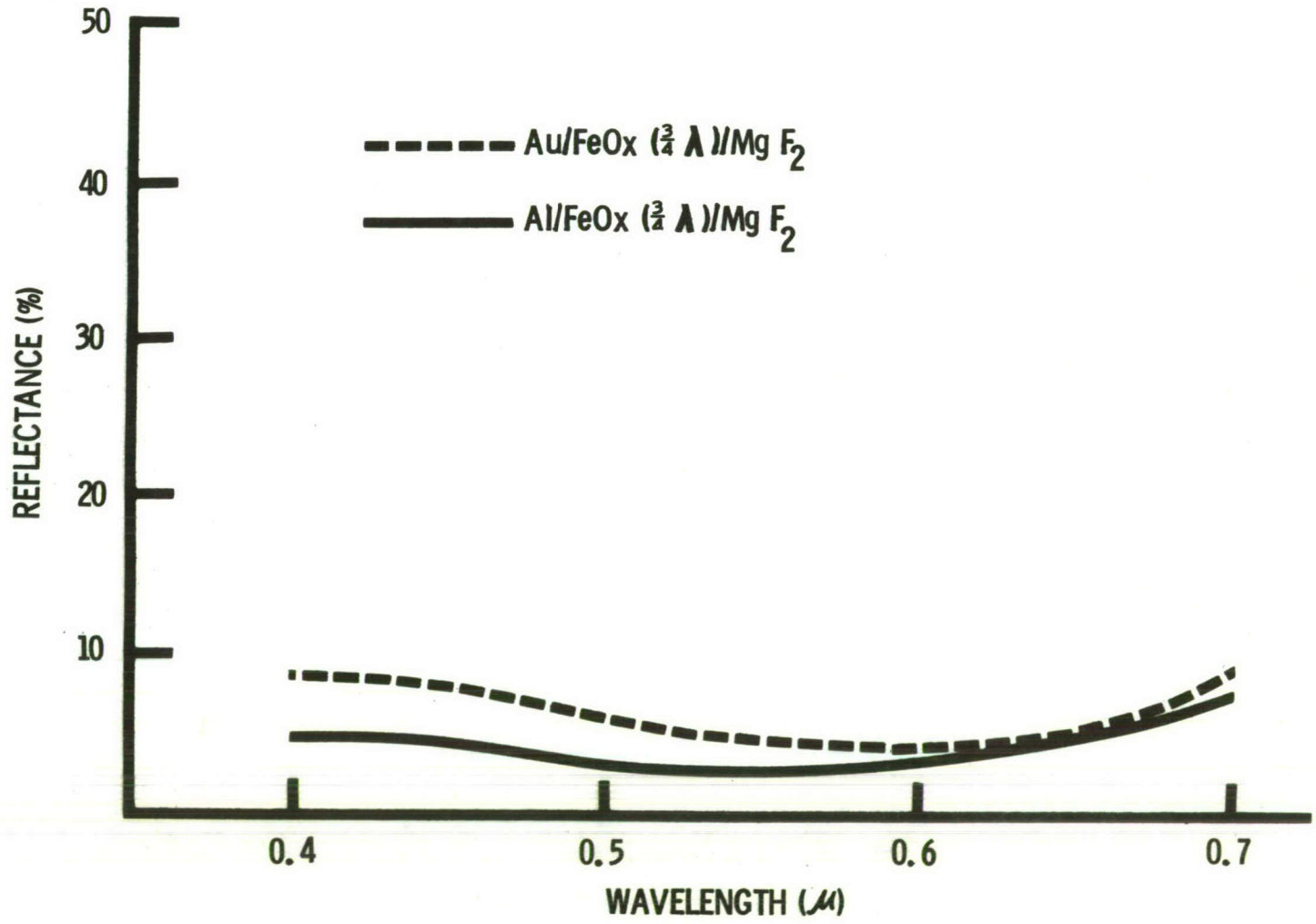


Fig. 10 REFLECTANCE vs. WAVELENGTH FOR $Al/FeOx(\frac{3}{4}\lambda)/MgF_2$ AND Al/SiO/Al/Si O

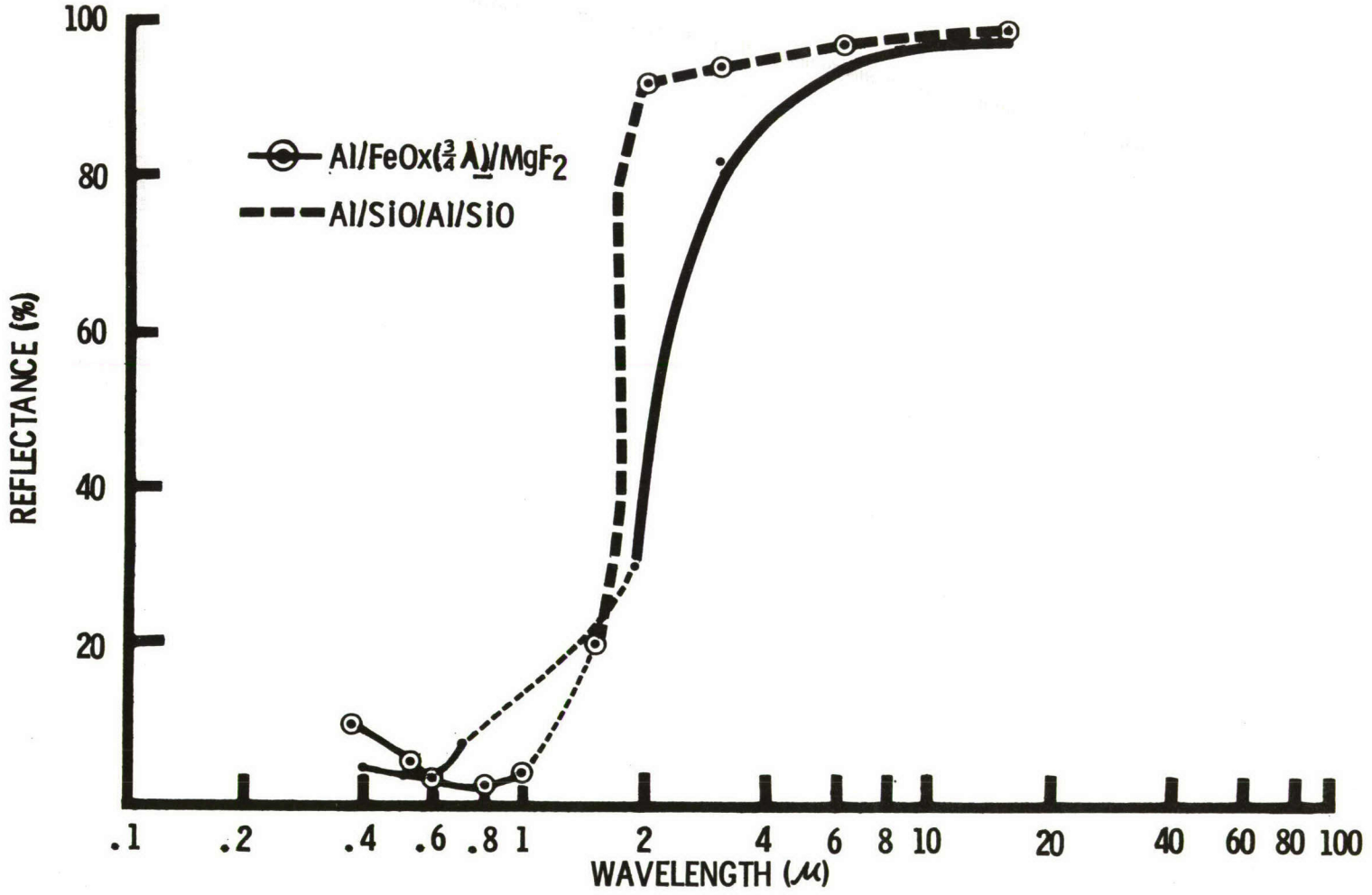


Fig. 11 SOLAR ABSORPTANCE (α) AND EMITTANCE (ϵ) vs. TEMPERATURE FOR Al/FeOx($\frac{3}{4}\lambda$)/MgF_2

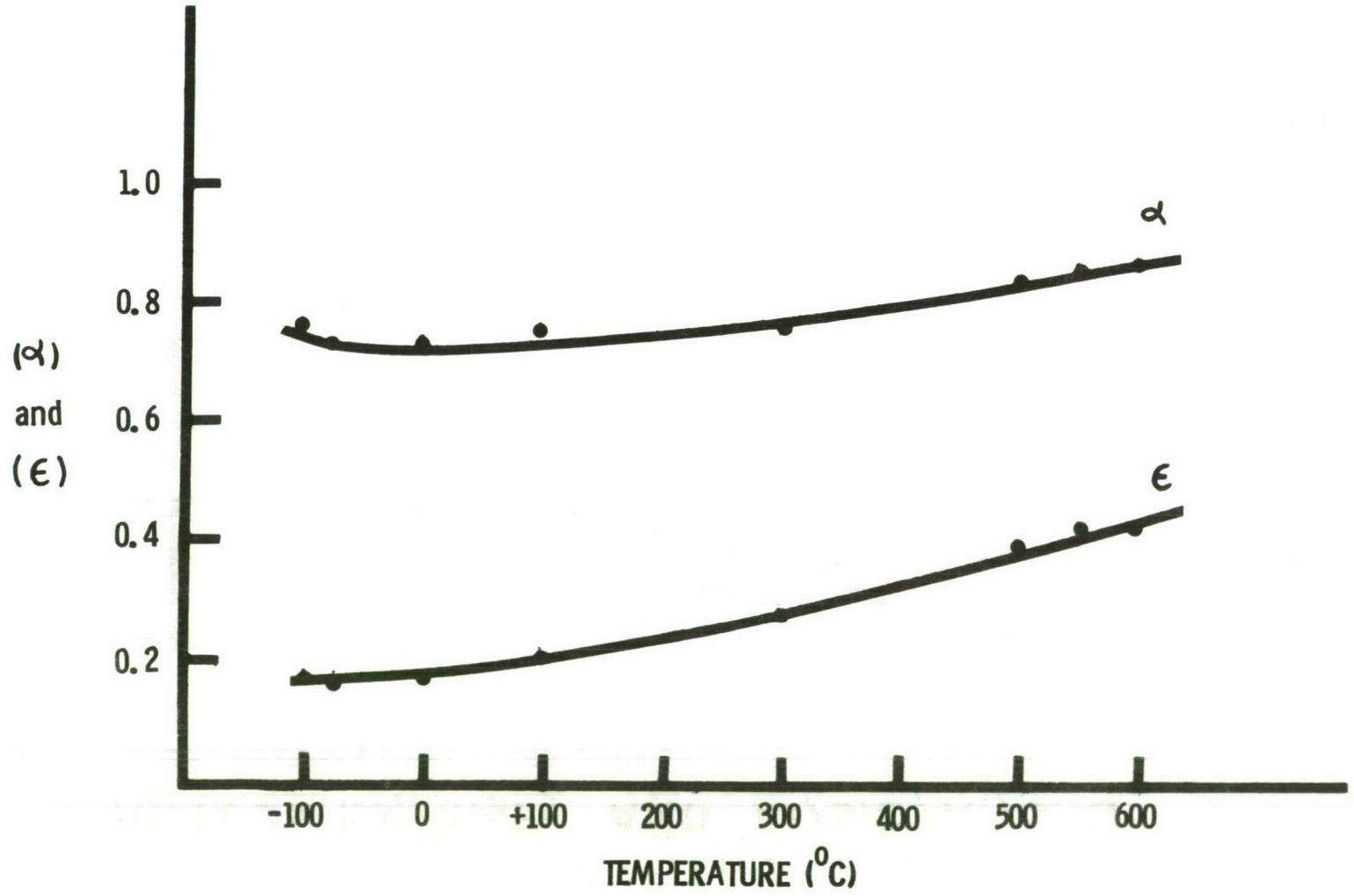
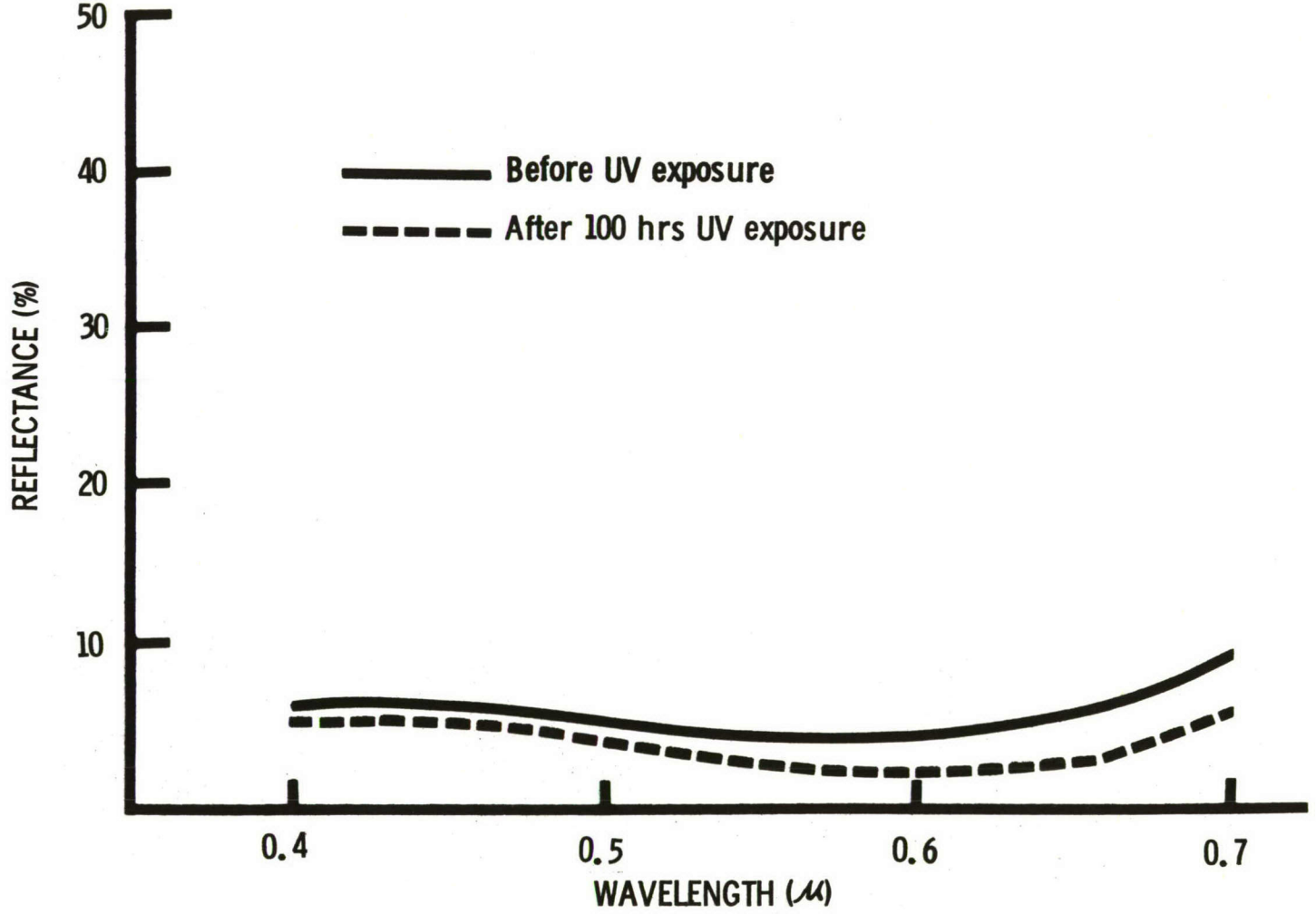


Fig. 12 REFLECTANCE vs. WAVELENGTH FOR Al/FeOx(¾λ)/MgF_2 BEFORE AND AFTER 100 HOURS VACUUM-ULTRAVIOLET EXPOSURE AT A PRESSURE OF 2×10^{-5} TORR



ENGINEERING SUPPORT

FOR

ELECTRIC PROPULSION SPACE TESTS

By

Peter D. Tannen
Captain, USAF

Space Vehicle Division (SWTSS)
Systems Engineering Directorate
Deputy for Test and Engineering
AIR FORCE SPECIAL WEAPONS CENTER
Kirtland Air Force Base, New Mexico

June 1964

ENGINEERING SUPPORT

FOR

ELECTRIC PROPULSION SPACE TESTS

1. INTRODUCTION:

1.1 Necessity of Flight Tests.

A number of electric propulsion devices have been built and ground tested for both the Air Force and NASA. Tests of these engines, particularly the ion or electrostatic type, must be conducted in large vacuum chambers. Although these chambers can achieve vacuums on the order of those found at 80-90 nautical miles, there is still some uncertainty on the validity of ground testing (Reference 1). Ion beam interaction with both the remaining atmosphere in the chamber and the chamber walls themselves can mask or render useless measurement of important performance parameters. Although better ground testing techniques are now becoming available, there still remains enough uncertainty to warrant flight testing of ion engines.

A successful flight test of an ion engine in space will relieve any doubts on the theory of charge neutralization and thrust build-up. Such a flight would also verify the results of extensive ground tests. Hopefully, only a very few flight tests would be necessary. Since ground tests are relatively inexpensive

and are repeatable, they are preferred over full-scale flight tests. If the flights confirm the ground test data, then no further flight tests would be needed.

1.2 Flight Test Program.

The office responsible for overall management of the Air Force Electric Propulsion Space Tests (Program 661A) is the Space Experiments Branch (SSTRX), Air Force Space Systems Division, Los Angeles, California. Sub-orbital ballistic flights were chosen as the most effective way to demonstrate thrust and charge neutralization of an ion engine. The ion engine used was developed and built by Electro-Optical Systems, Inc., Pasadena, California. Because of budgetary limitations and because of in-house capability to design, fabricate, and test the support portion of the payload was available at the Air Force Special Weapons Center, the entire flight package was not contracted out. The group from AFSWC who accomplished design of the first payload had considerable experience in the design of space payloads. Design, fabrication, testing, and first flight of the payload were completed within one year of the go-ahead.

Also available at AFSWC was a group of Air Force engineers experienced in booster development and flight planning. This group, Space Vehicle Division, provided the basic data on the vehicles which were then available, and aided SSD in choosing the best

booster. After the first flight, a re-organization at AFSWC placed the Space Vehicle Division in charge of developing two additional support payloads for Program 661A. This division continued to provide the various documentation which is needed to launch a space probe. This includes not only the basic Range documents, such as the Program Requirements Documents and Range Safety Report, but also the Vehicle Assembly Manual, Pad Procedures, and the Countdown Manual. The Space Vehicle Division insures overall flight effectiveness by providing payload/booster integration and by aiding the launch crew in vehicle assembly, checkout, and launch. The division also provides preflight and post-flight analyses of all flights. This paper primarily describes the two payloads developed by the Space Vehicle Division.

2. BOOSTER VEHICLE:

2.1 General Description.

The booster for Program 661A flights is the Standard Launch Vehicle I-B (SLV-IB in its modified three-stage version; Figure 1). The interstage structures, heatshield, wiring, and fins were designed and built by Ling-Temco-Vought, Inc. LTV also provided aerodynamic data and launch support. The vehicle is unguided, but is aerodynamically stabilized by fins and spin in the sensible atmosphere, and by spin alone for the remainder of its powered flight. Spin is imparted to the vehicle by two tangentially-mounted rockets which

are fired after the booster clears the launcher. These rockets fall away shortly after firing and spin is maintained by canted tips on the first stage fins and by canting of the second stage fins. The final spin rate is 2.7 revolutions-per-second.

2.2 Stages.

The following rocket engines are used on the SLV-IB (modified):

1st stage:	Thiokol XM-33E7 "Castor"
2nd stage:	ABL X259-A4 "Antares"
3rd stage:	Aerojet AJ-10-41 (30KS8000) "Alcor"
Spin motor:	Aerojet .24KS-9900 "Hot Seat"

The first stage is fired from the launch van; the second and third stages are fired by the stage immediately preceding them. That is, when the first stage chamber pressure falls below 50 psi, a pressure switch fires the second stage. The third stage is fired similarly, except there is a short delay between second stage pressure drop and third stage ignition. The heatshield is separated from the vehicle during this delay period. All stages are statically balanced and the third stage and payload are dynamically balanced.

2.3 Performance.

The nominal performance of this vehicle is:

Launch angle: 84° elevation, 290° azimuth
(West Coast launch)

Payload weight:	250 pounds
Apogee Altitude:	1,200 NM
Time above 200 NM:	29 minutes
Range at re-entry:	1,450 NM

3. PAYLOAD DESCRIPTION:

3.1 Ion Engine.

The test engine for all Program 661A flights is the Electro-Optical Systems cesium contact ion engine developed for the Air Force Aero-Propulsion Laboratory, Wright-Patterson AFB (Reference 2). This engine was designed specifically for the Electric Propulsion Space Flight Tests. The EOS portion of the payload consists of two basic subsystems. The largest of these is the electronics section, which contains the power supplies, signal conditioning equipment and programmer. The other main subsystem is the thruster itself, which includes the propellant supply and feed system, the ionizer, the acceleration and beam-forming electrodes, and the beam neutralizer (Figure 2). This engine has a thrust capability of several millipounds and a specific impulse of several thousand seconds.

3.2 AFSWC Support Package.

The basic purpose of the AFSWC portion of the payload is to provide system support for the electric propulsion space tests. This portion of the payload contains engine performance sensors, a

telemetry system, an attitude stabilization system, and a power system (Figures 3 - 7).

3.2.1 Engine Performance Sensors.

Two basic physical measurements of engine performance are made by the support payload; these are thrust and charge neutralization. Thrust is measured with a Honeywell "micro-G" accelerometer developed for AFSWC especially for use on this project (Reference 3). This instrument is a highly damped, pendulous linear accelerometer which is centered by means of a DC torquer. The rebalance torquer current is proportional to the applied acceleration. A voltage proportional to this current is telemetered. The accelerometer has a threshold of 1×10^{-6} "G" and a usable range up to 20×10^{-6} "G". The pendulum assembly is pivoted on jewel bearings which are dithered to reduce friction and null errors. A temperature control circuit and internal heater maintain the accelerometer at a constant temperature. This accelerometer is mounted very close to the payload center-of-gravity to minimize effect of rotational accelerations.

Space Technology Laboratories' electric field strength meters measure the floating potential of the payload with reference to free space (Reference 4). If the ion beam is not completely neutralized, then the payload will build up a large negative electrostatic charge. Data on this charge build-up and its

effect on thrust are vital to the further development of ion thrustors. The STL sensors developed for this project are essentially open triodes with the payload skin acting as the grid. If the payload is at zero volts potential with respect to free space, the cathode and anode currents are approximately equal. A positive charge build-up causes both the cathode and anode currents to fall equally (Figure 8). However, a negative charge on the payload drives emitted electrons into space causing the cathode current to increase while the anode current decreases. As shown in Figure 8, the current separation is biased to start at a slight positive payload potential. Two E-meters are carried on each payload; one directly under the beam, the other facing 180° from the beam. This arrangement of sensors will make it possible to check on local variations in charge due to the ion beam. The cathode and anode currents of each sensor are simultaneously transmitted. Neither the micro-G accelerometer nor the E-meters were modified after the first launch.

3.2.2 Attitude Stabilization System.

Although the micro-G accelerometer is mounted very close to the center of gravity of the payload, slight misalignment coupled with small rotation rates could produce centripetal accelerations which would swamp the readings of motor thrust. In order to prevent this from occurring, a three-axis attitude stabilization system was incorporated into the payload. Gas reaction jets were

not considered feasible, since any gas atmosphere around the payload would cause deterioration of the engine. Also, the thrust from pulsed gas systems could easily mask the engine thrust measurements. Reaction wheels were considered, but the best solution in terms of weight, volume and power was the Ling-Temco-Vought twin-gyro controller (Reference 5). This system uses gyroscopic output torques to stabilize and control the payload. Two single-degree-of-freedom gyros control each axis (Figure 9). Because their spin directions are opposite and they are torqued in opposite directions, the twin-gyro combination will exchange momentum in one axis only (Figure 10). Each axis has a total momentum capability of approximately eight ft-lbs-sec. Tests of the twin-gyro system have shown that it is capable of positioning a space payload to an attitude error on the order of 1 arc-second.

Error signals for the controller are provided by two 2-degree-of-freedom gyros which make up the attitude reference system (Reference 6). These gyros are mounted on a servo-driven, roll-stabilized platform. This unique feature allows the system to be used not only as an attitude reference for closed-loop control, but also to provide data during the boost phase of flight. The roll-stabilized reference system was conceived and developed by the Whittaker Corporation. The basic design was modified to meet special AFSWC requirements. The most important of these was the need to replace the acceleration switch used on the first flight to provide

the separation signal. Since the controller is momentum-limited, the payload must be despun and separated from the empty third stage before the attitude control system is activated.

Three small retro-spin rockets (Figure 3) are fired after third stage burnout. These have the capability of despinning the entire stage and payload from 3 rps to 0 rps in 1 second. Since the vehicle is spinning at approximately 2.7 rps at the time of burnout, they have a slight retro-spin capability. The platform servo motor on the Whittaker system is actually a servo-tachometer. When the tachometer output falls below a certain level, it actuates a trigger circuit, which in turn fires an explosive bolt separation mechanism. Tests have shown that separation will occur at rates below 10 degrees/second. The system has been tested under conditions which simulate the payload deceleration.

After separation, the reference system is recaged, then uncaged, to provide space-oriented reference axes for the remainder of the flight. If this were not done, the payload would be referenced to the prelaunch conditions and large control error signals would accumulate.

3.2.3 Telemetry System.

The Program 661A payload uses two 10-watt FM telemetry transmitters manufactured by Advanced Electronics Corporation. Both of these telemetry links carry all the primary experimental data, so that loss of one transmitter will not jeopardize the

data recovery. Seven FM subcarriers are used on each link. Six of these carry continuous analog data, while the seventh (IRIG Band E) receives its data from a 45-segment, 20 frame-per-second PAM commutator. This commutator samples the engine performance data from the EOS portion of the payload and monitors various functions in the AFSWC portion. The micro-G accelerometer, E-meter, and commutated data are carried over both telemetry transmitters, while the attitude reference and control system data are split between the two links (Figure 4).

One of the two transmitters is switched over during the boost phase to carry booster data. This transmitter is switched back to experimental data at payload separation. The following booster parameters are measured: longitudinal and lateral acceleration, vibration, second and third stage chamber pressures, spin rate, pyrotechnic ignition, and stagnation point heating. Most of the instrumentation for these functions is mounted on a tray in the payload/booster interstage (Figures 3 and 11). The Whittaker attitude reference system provides data on the vehicle attitude. The accelerometers, which are made by Pace Engineering, contain their own subcarrier oscillators, so the booster telemetry link carries a total of nine channels during boost phase. Statham pressure transducers provide the chamber pressure data, while spin rate is measured by a Schoenstedt Instrument Co. magnetic

aspect sensor. The output of this instrument is a sine wave whose frequency is coincident with the spin rate of the vehicle. Stage tip-off and vehicle stability information can be derived from the magnetic aspect sensor.

Pyrotechnic ignition is sensed by ignition monitors, which were designed and built at AFSWC. These instruments transform the ignition current pulse into a voltage pulse compatible with the payload telemetry. A Microdot thermistor mounted at the top of the EOS payload provides data on stagnation point heating. Other items monitored by the telemetry are payload internal temperature, power supply voltages, and the reference system cage-uncage cycle.

3.2.4 Power System.

Three separate Yardney silver-zinc battery systems provide power for the payload. The largest of these are two packs of PM-16 cells which provide 25 amp-hr, 56V DC power to the ion engine. Power converters within the EOS electronics section condition this DC power to the requirements of the engine. Two PM-5 battery packs provide 28V, 15 amp-hr, DC power to the AFSWC payload. An Engineered Magnetics DC-to-400 cps AC static inverter produces the power required for the attitude control gyros and their electronics. Power for the Whittaker gyros is provided by that system's own inverter. Various other regulated power supplies

are used for the transmitter plate voltage, the micro-G accelerometer, thermistors, and calibration voltages. Power for the E-meter filaments is provided by a pair of PM-1 battery packs.

3.2.5 Systems Integration.

Although the major subsystems were procured from experienced contractors, there still remained the difficult task of combining these systems into a working payload. Numerous problems were anticipated, met, and overcome. Small volume due to heatshield restrictions, dynamic balance requirements, and ease of parts replacement were three conflicting restraints in the physical packaging of the payload. The two later payloads have a density of approximately 70 lbs/ft^3 and have better accessibility than the first payload while maintaining its inherently good balance.

Various control functions had to be added to insure proper systems operation. A pair of mechanical timers provide redundant signals to activate such critical functions as the ion engine programmer, payload despin, control system activation, and separation backup. These timers are started at heatshield separation. Since there is some delay before the control gyros are fully erected, the reference gyros must remain caged for this delay period. If this were not done, extremely large error signals could build up and swamp the controller. The payloads were designed so the entire telemetry system can be calibrated up to two minutes before launch.

Also, all important "Go-No-Go" parameters can be monitored prior to launch. These control functions had to be designed so that no unwanted side effects could occur due to switching transients or noise. Any possibility of intersystem electrical interference was eliminated by careful design and testing.

4. FLIGHT TESTS:

4.1 Test Code A (SWC-16):

During the second quarter of Fiscal Year 1963, the Air Force made the first attempt of testing an ion engine in space. The booster requirements were to place a 250-pound payload above 200 nautical miles for at least twenty minutes. The actual flight time above 200 nautical miles was over twenty-five minutes (Reference 7). Telemetry was received for over thirty minutes, and excellent booster performance data were obtained. However, when the ion engine's high voltage was turned on, there was an extreme current drain and a corresponding drop in primary battery pack voltage. At the same time, there were indications of intermittent arcing on one of the high voltage telemetry traces. There was no indication of engine thrust build-up. Signal strength data indicated that the payload separated from the booster with a spin rate approximately six times that expected. The attitude control system's telemetry outputs indicated that the roll gyros did not erect and that there was an oscillation in yaw (Reference 8).

4.2 Causes of Failure.

The most probable cause of the ion engine failure was determined to be relatively high gas densities within the high voltage section of the EOS electronics. It is very likely that gases escaping from the AFSWC portion of the payload, particularly the batteries, accumulated in the electronics. Although venting was provided, it was not sufficient to keep the pressure within the high voltage section low enough. Since then, the EOS electronics have been hermetically-sealed and the AFSWC package has been adequately vented. Vacuum tests have shown that these fixes are satisfactory (Reference 9).

Control system tests at AFSWC determined the cause of controller failure (these tests are covered in Section 5.3). Payload separation was unsatisfactory on the first flight due to a basic system fault. An acceleration switch was used on this flight to sense despin and fire the separation mechanism. A study conducted at AFSWC showed that this switch could not perform adequately under the deceleration rates and vibration present during despin. Tests also showed that the process of dynamically balancing the vehicle and payload could have altered the characteristics of the acceleration switch due to the repeated spin-up and despin associated with the balancing operation.

5. TESTING:

5.1 Test Philosophy.

The basic philosophy behind all testing of the second and third payloads was that the payload should be in flight configuration and operating during tests. This philosophy was followed whenever equipment limitations allowed. Two notable exceptions were that the E-meters should not be turned on at pressure altitudes below 80 miles and the micro-G accelerometer is normally "swamped" in a 1G field. For all tests, the AFSWC payload was operating on full internal power and transmitted data were displayed in real time. The first payload was not fully tested due to the pressure of meeting a flight schedule. Proper testing would probably have prevented the failure of this flight test. The following tests were performed after the first flight.

5.2 Environmental Tests.

The AFSWC payload was environmentally tested in longitudinal shock, lateral and longitudinal vibration, temperature, and altitude. During the vibration and shock testing, the booster instrumentation was telemetered and recorded. These vibration and shock levels closely duplicate those expected on the vehicle during boost phase. The highest vibration levels occur during second stage burning, while the largest shock occurs at third stage ignition. The shock, vibration, and temperature tests were accomplished at AFSWC as part of the normal workload under the Deputy for Test and Engineering. No special equipment had to be purchased or developed since the in-house equipment was more than adequate to meet the needs of this

payload. Altitude tests of the payload up to 130,000 feet can also be accomplished at AFSWC and individual components can be tested beyond 200 miles altitude. Complete vacuum tests were accomplished at both Arnold Engineering Development Center and Electro-Optical Systems, Inc., so these tests were not duplicated at AFSWC.

5.2.1 Shock Test.

The payload and interstage were subjected to the following shock load: 30g - peak, half square wave of 11 millisecond (ms) duration. The payload systems operated satisfactorily before, during, and after this test; the reference system drift was less than one-half degree.

5.2.2 Vibration Tests.

The following vibration levels were applied to the payload:

Longitudinal:	0.01 in double amplitude 20-100 cps, 10 g's 100 cps - 2000 cps, 30 second total sweep.
Lateral:	0.01 in double amplitude 20-100 cps, 5 g's 100 cps - 2000 cps, 30 second sweep.

The payload functioned perfectly during these tests. The vibrometer output showed that the fiberglass interstage provides considerable damping and attenuation above 100 cps. The lateral vibration tests were conducted in both the pitch and yaw planes. The lateral accelerometer was able to follow these vibrations to

frequencies as high as 40 cps. Vibration tests of a payload structure were also done to check for resonance points. The structure was weighted to simulate the flight loading and vibration pickups were attached to a number of points on the structure. No significant resonances were noted.

5.2.3 Temperature Test.

The payload was flown through a simulated launch temperature environment. This included one day at 50°F and 90% humidity, followed by a 70°F dry air purge for two hours, simulating the prelaunch conditions. The launch conditions were fast rise from 50°F to 125°F, and hold 125°F for two minutes, drop chamber temperature to -65°F or less, and hold for thirty minutes. All systems were operating during the test, and telemetry indicated that they operated normally.

5.2.4 Altitude Tests.

Altitude tests were conducted at both EOS and AEDC. The tests at AEDC closely simulated the flight profile of the rocket vehicle (Reference 10). The ion engine was turned on as programmed on an actual flight and performed as expected. There was no significant arcing. The series of tests conducted at AEDC were designed to prove the modifications made to the payload after the first flight test failure (Reference 9). The tests performed at EOS were slow pump-down and designed to check the AFSWC/EOS compatibility and qualify the flight package. The payload operated satisfactorily on

internal power throughout all altitude tests. Similar tests at EOS were performed on the first payload, but significantly, the batteries were outside the test chamber. These tests constituted the most severe temperature test of the payload since heat transfer was by radiation alone.

5.3 Control System Tests.

In order to confirm complete compatibility of the Whittaker reference system, the LTV controller, and the payload, closed-loop control system tests were performed. These tests were conducted on an AFSWC-built air-bearing table (Figure 12). This table is a redesigned version of one built by Ling-Temco-Vought. The table and payload are completely supported by an air column which allows virtually friction-free motion. It has full 360° freedom in roll, but is limited to $\pm 30^{\circ}$ in pitch and yaw. This piece of equipment has proved invaluable in testing the control system. When the table is properly balanced, the controller can easily detect mass shifts of less than one gram.

Tests conducted at AFSWC on the second payload showed that although the controller can stabilize the payload against applied torques, it could not remove vehicle rates above 3 degrees/second. This was due to the back precession ($\omega \times H$) torques felt by each individual gyro. Although the net torque on the payload is zero, each gyro must be rotated against its precession torque to control the payload. The torque motors had insufficient output to do this. Proper testing early in the program would have revealed this design fault. Neither the contractor nor any of the engineers associated

with the project suspected that the controller had this basic fault. Previous tests conducted by the contractor were torque tests only and did not have initial rate errors. Also, analytical studies by both LTV and Aerospace Corporation failed to reveal the discrepancy. The controller was fixed by adding reduction gearing to the motor to provide the necessary torque, and by changing loop electronic gains for system stability. The payload moments of inertia were measured on the AFSWC torsion pendulum (Figure 13). These moments had to be accurately measured to predict the positioning of the despin rockets and the performance of the control system.

6. SUMMARY:

The first flight test of an ion engine was unsuccessful due to a failure in the engine. Sufficient data were received to identify the failure as high voltage arcing within the engine electronics; various systems within the AFSWC support package also did not function as expected. A contributing factor in these failures was the lack of adequate testing. This was generally caused by the compressed time scale between the start of the program and the first flight.

Numerous tests have been conducted on the payloads since the first flight. These tests have proven that the modifications made on the payloads are sufficient to prevent recurrence of the first flight failures. The ability of the payload to operate successfully through these tests has greatly increased the confidence level of the payload.

ACKNOWLEDGEMENTS

The author acknowledges the cooperation of Space Experiments Branch (SSD); Electro-Optical Systems, Inc.; and the major contractors, Honeywell Corporation, Ling-Temco-Vought, Inc., Space Technology Laboratories, and Whittaker Corporation, in the development of these payloads. He thanks the Deputy for Test and Engineering (AFSWC) for the excellent engineering and fabrication support rendered throughout the project. The author also expresses special thanks to Lt Charles H. Radoy, Project Electrical and Telemetry Engineer, for his assistance in the preparation of this paper.

A five-minute film will be presented with the paper.

REFERENCES

1. Baldwin, Dr. G. C., et. al., Neutralization of a 30 KW Ion Engine Array, G. E. Report R63FPD130, General Electric, Cincinnati, Ohio, March 1963 (Unclas Report).
2. Winsor, K. W., Densmore, J. E., and Davis, J., (U) Ion Thrustor Program, ASD TDR 63-848, Aeronautical Systems Division, Wright-Patterson Air Force Base, Ohio, March 1963 (Conf Report).
3. Rasmussen, W., et. al., (U) Micro-G Accelerometer System, MH Aero Report 2888-TR1, Minneapolis-Honeywell Regulator Co., Minnesota, Final Report, Contract AF 29(601)-5383, October 1962 (Conf Report).
4. Forbes, S. G., Fabrication and Delivery of Electric Field Strength Meters, 8745-6005-RU-000, Space Technology Labs., California, Final Report, Contract AF 29(601)-5357, September 1962 (Unclas Report).
5. Schellhammer, A., Twin-Gyro Attitude Controller, LTV Report R24-553, Ling-Temco-Vought, Inc., Texas, Final Report for Contract AF 29(601)-4963 (Unclas Report).
6. Dugan, C., Miniature Attitude Reference System, Whittaker Controls and Guidance, Report FE 8477A, Proposal for Contract AF 29(601)-6245, July 1963 (Unclas Report).
7. Space Vehicle Branch, (U) Program 661A Vehicle, Test Code A, BSJ-16-8, AFSWC, Kirtland AFB, New Mexico, April 1963 (Conf Rept).

REFERENCES (CONT'D)

8. Moulds, W. J., Jr., (U) Final Report for the SWC-16 Payload, AFSWC TDR 63-30, AFSWC, Kirtland AFB, New Mexico, April 1963 (Conf Report).

9. Sitney, Dr. L. R., Shukay, Capt T. R., High Altitude High Voltage Breakdown in an Electric Propulsion Flight Test System. Paper presented at AGARD Conference, Marseille, France, April 1963 (Unclas Report).

10. Latture, N. C., and Cheesman, C. E., Flight-Profile Test of the Space Systems Division Ion Engine Flight Test Package in the 7-Foot Aerospace Research Chamber, ARO, Inc., Arnold Engineering Development Center, Tennessee, June 1963 (Unclas Report).

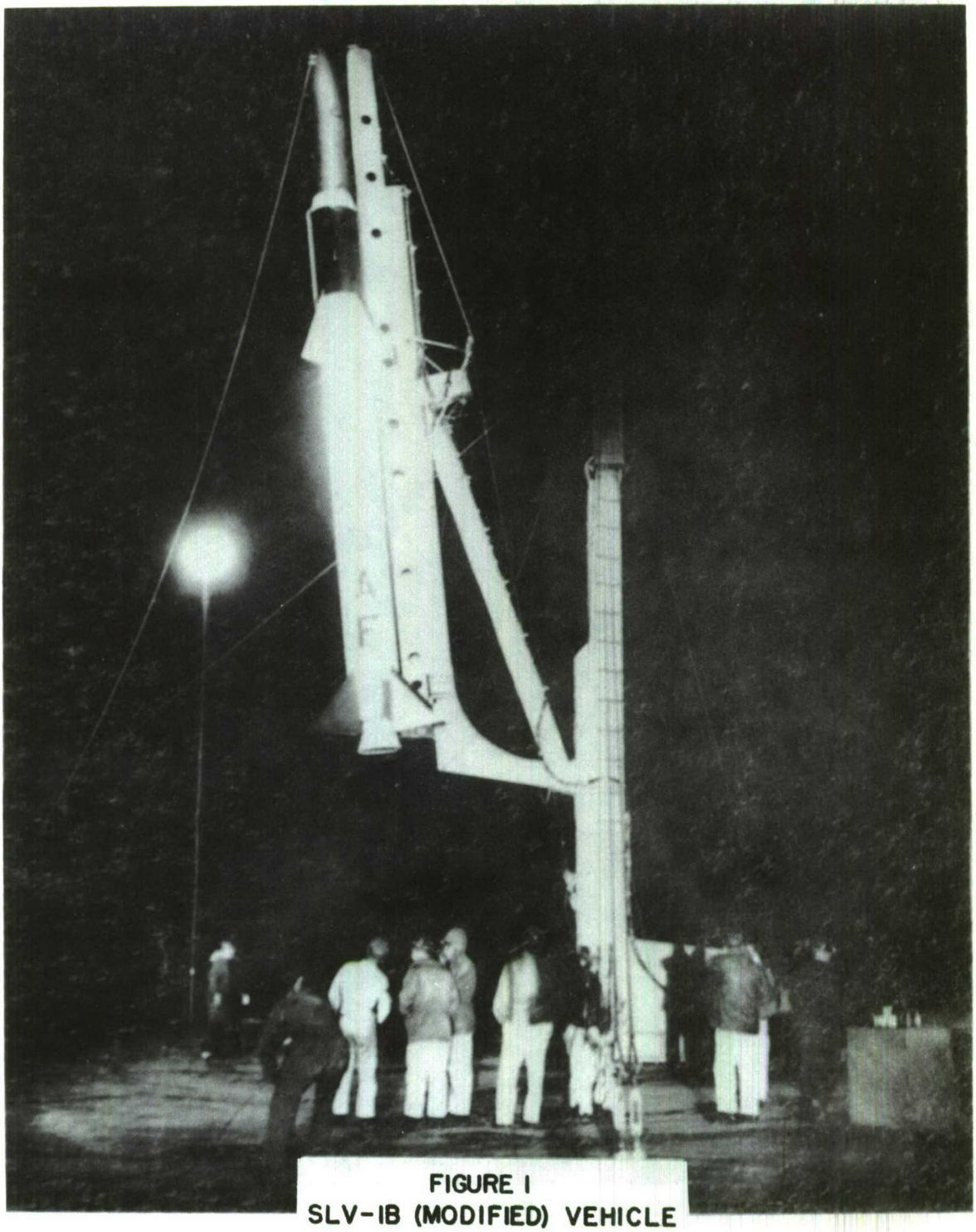


FIGURE 1
SLV-1B (MODIFIED) VEHICLE

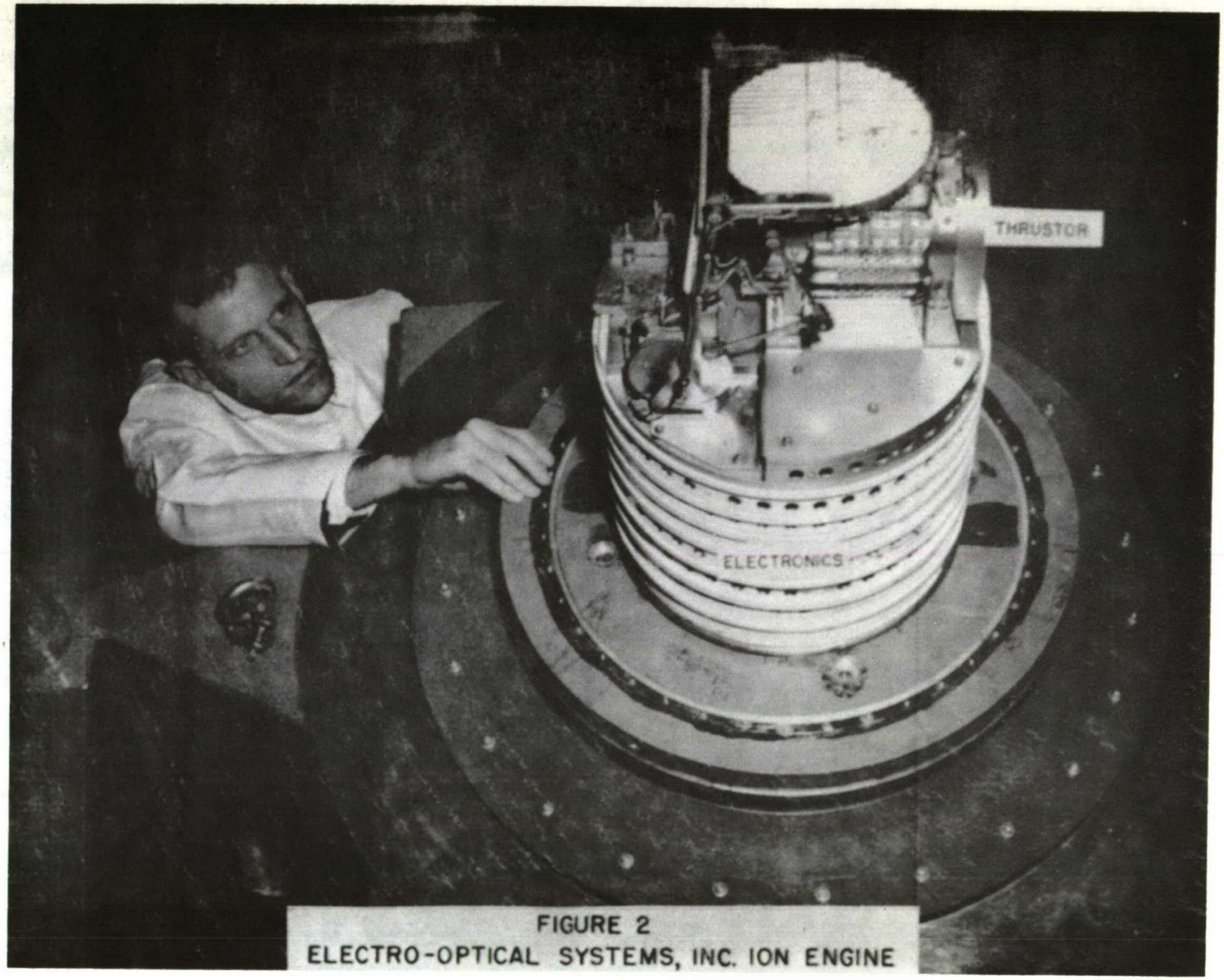


FIGURE 2
ELECTRO-OPTICAL SYSTEMS, INC. ION ENGINE

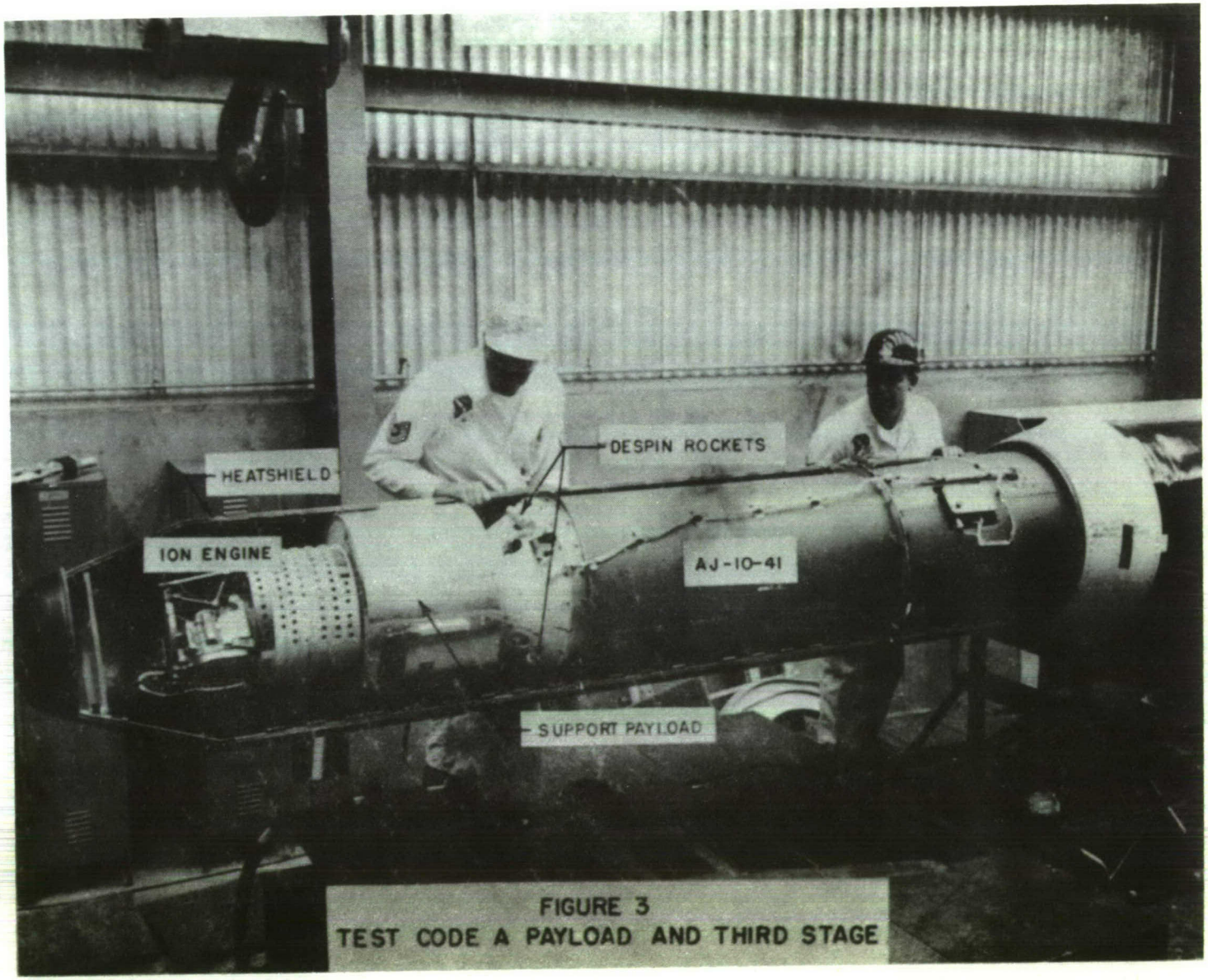


FIGURE 3
TEST CODE A PAYLOAD AND THIRD STAGE

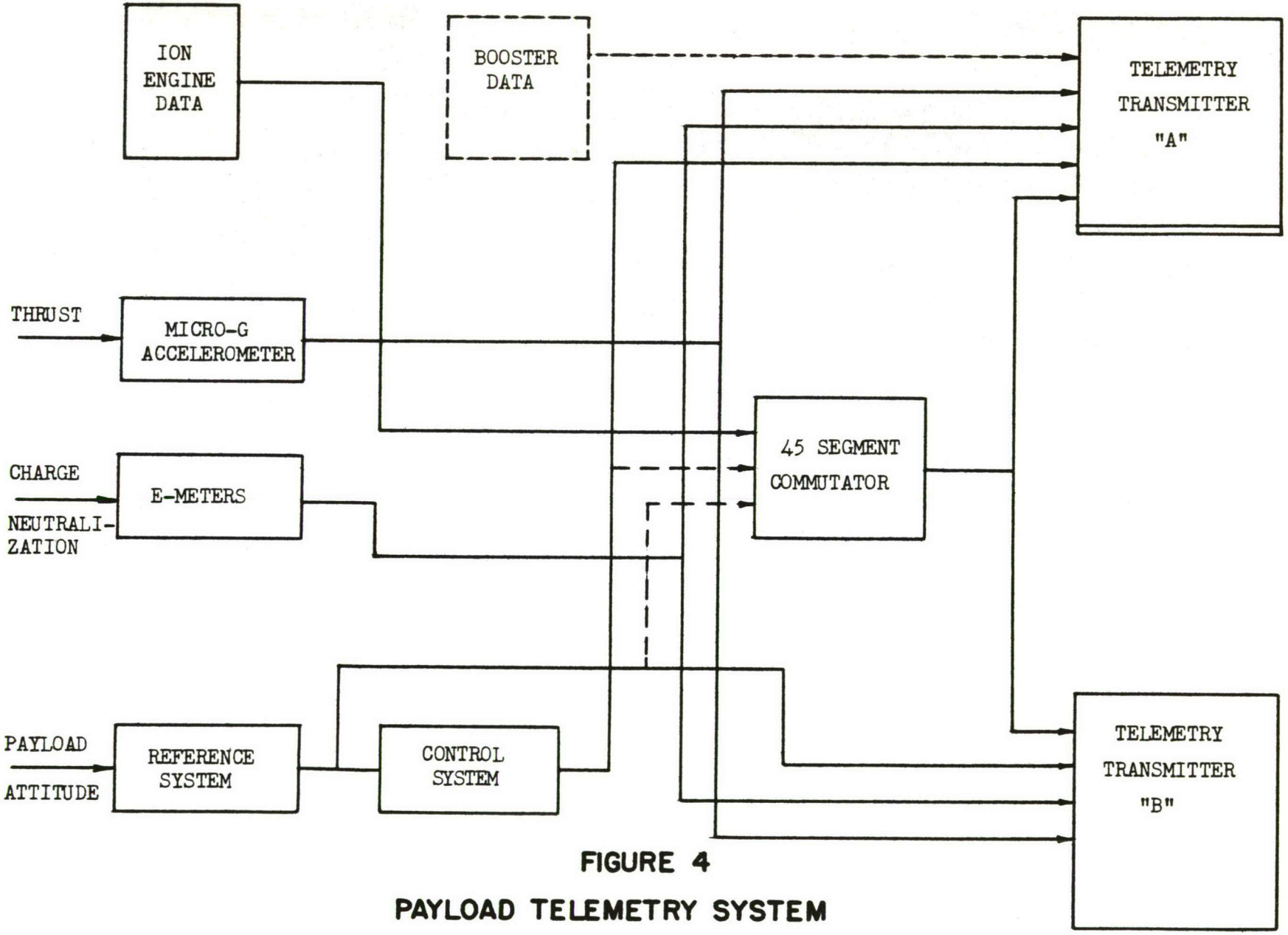


FIGURE 4

PAYLOAD TELEMETRY SYSTEM

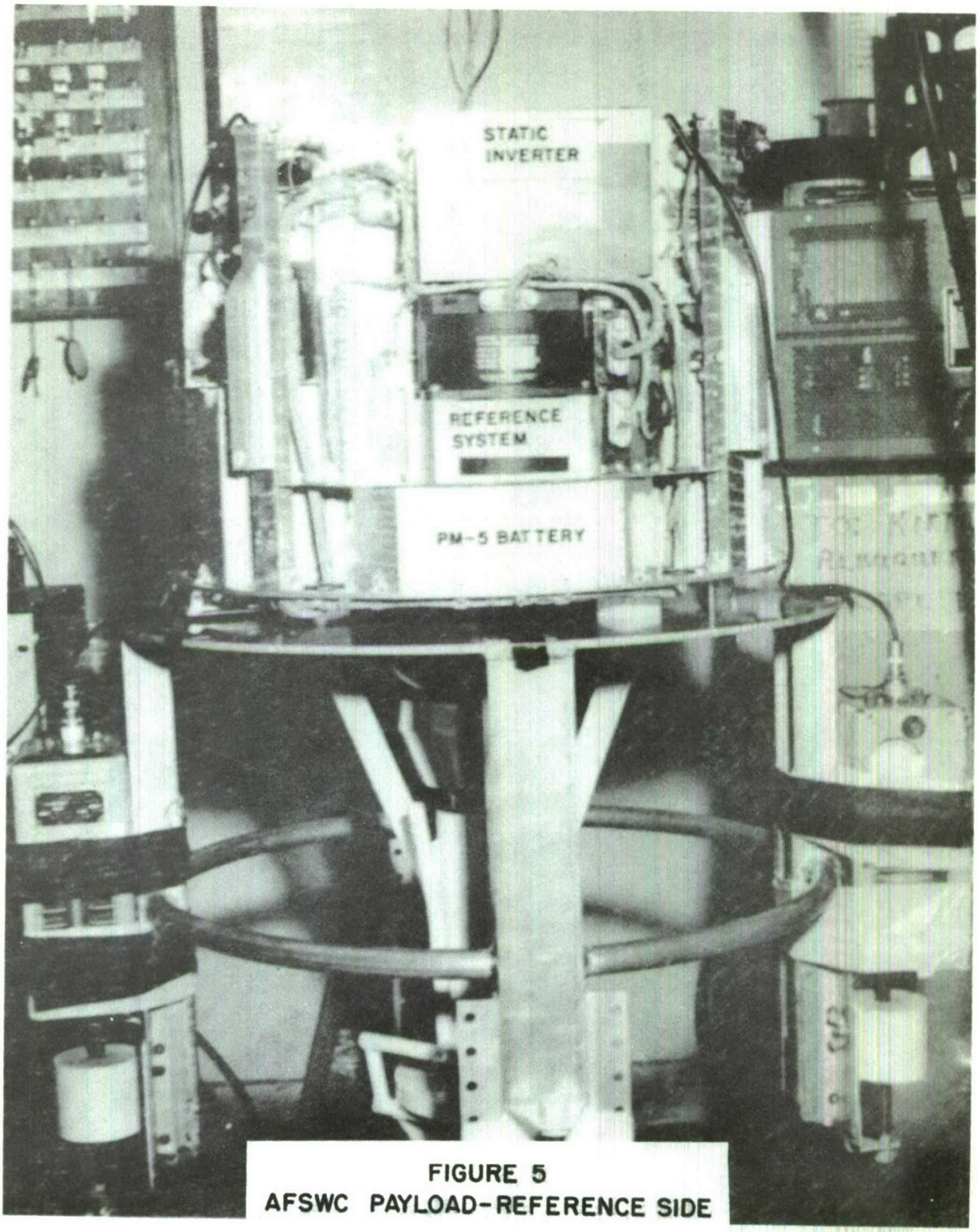


FIGURE 5
AFSWC PAYLOAD-REFERENCE SIDE

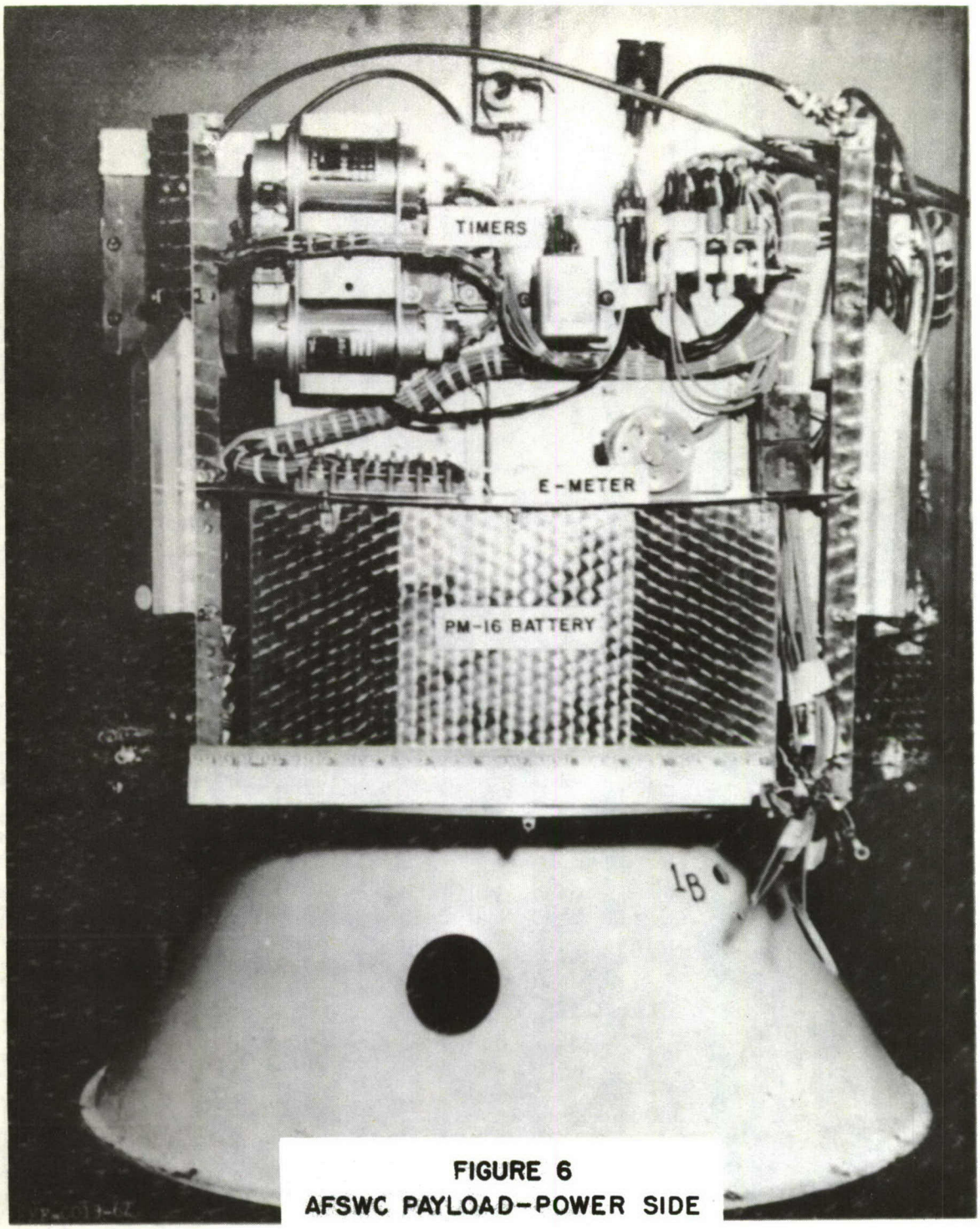


FIGURE 6
AFSWC PAYLOAD–POWER SIDE

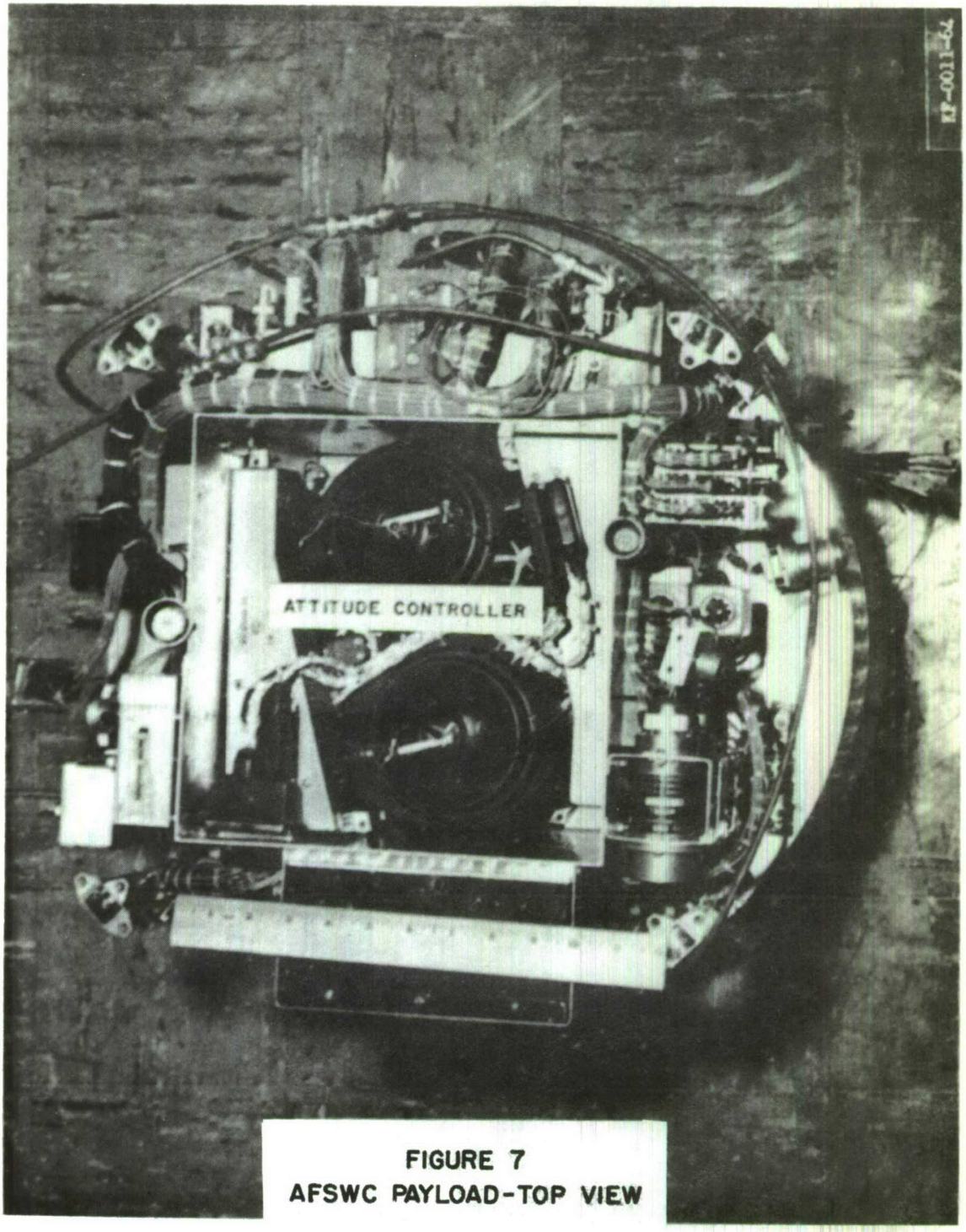


FIGURE 7
AFSWC PAYLOAD-TOP VIEW

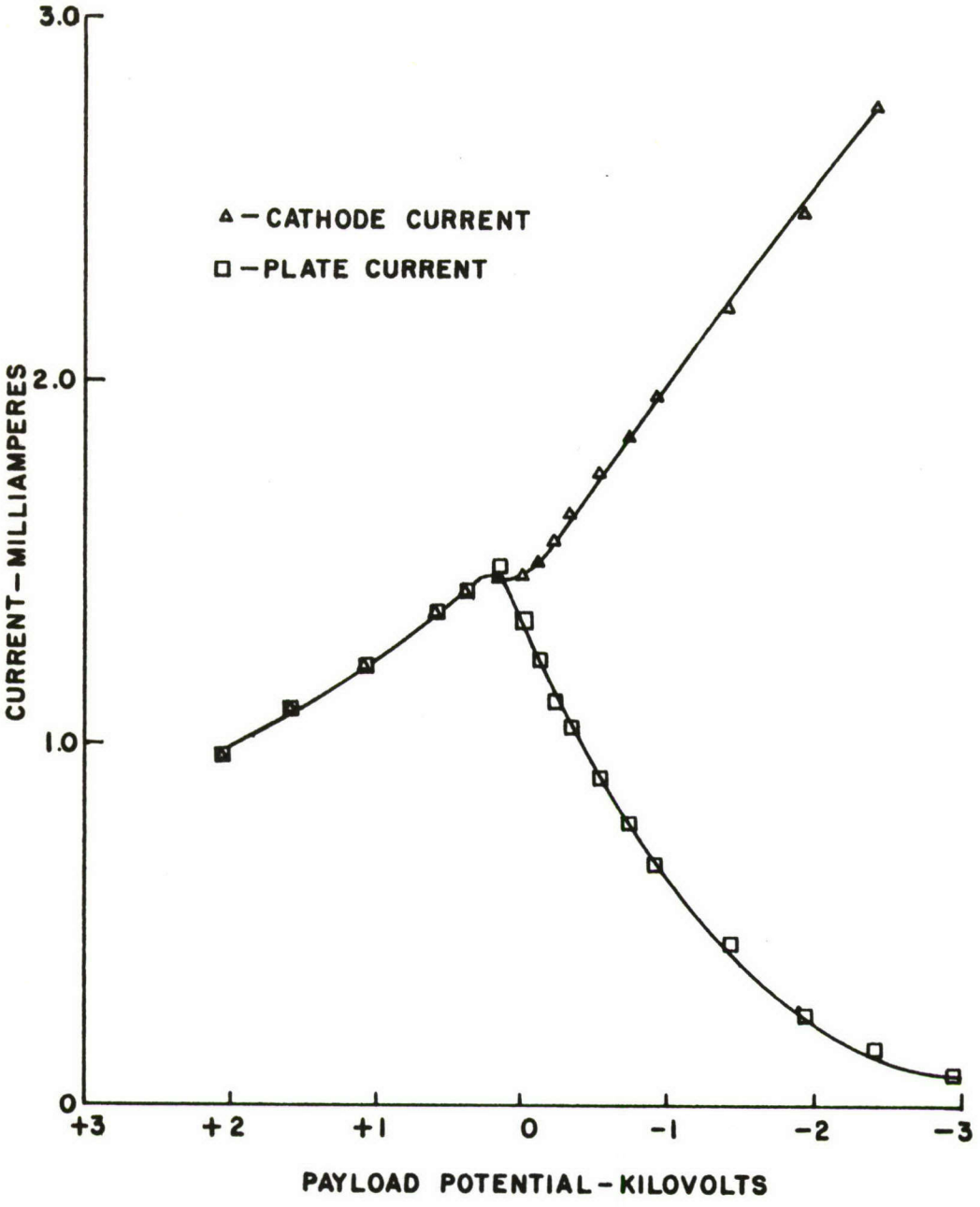
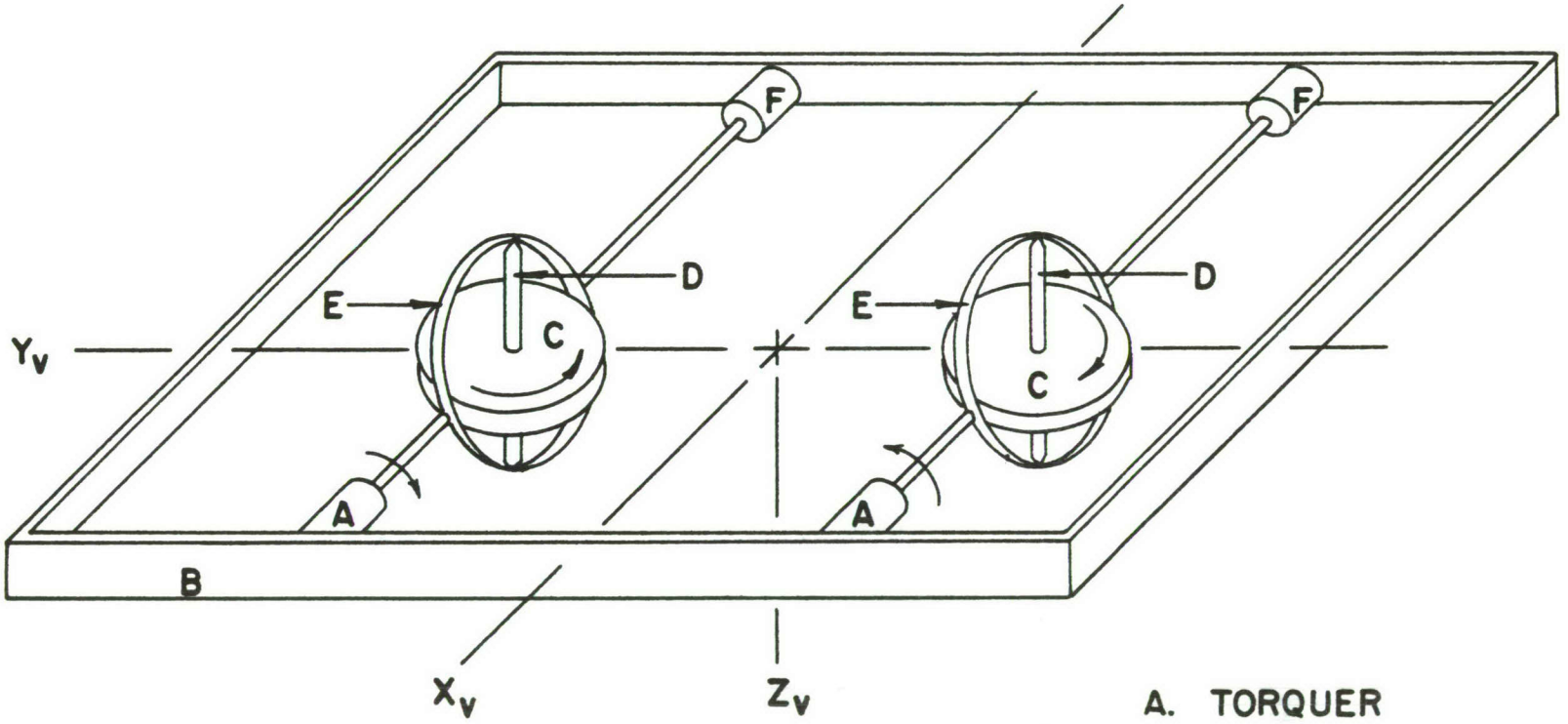


FIGURE 8

E–METER OPERATING CHARACTERISTICS

(REF. 6, FIG. 17)

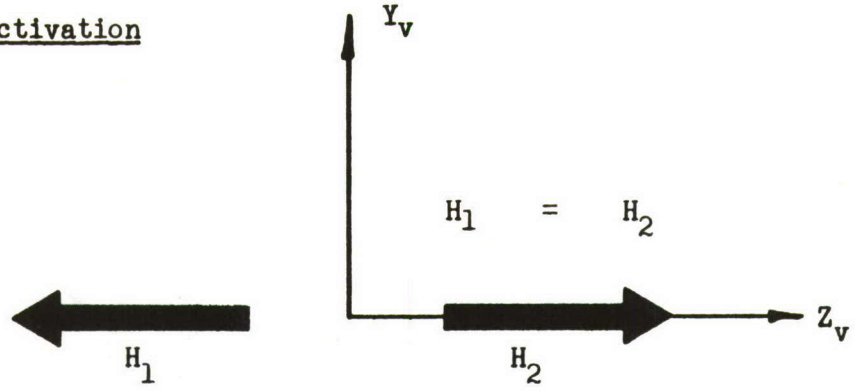


A. TORQUER
B. VEHICLE FRAME
C. GYRO ROTOR
D. GYRO SPIN AXIS
E. GIMBAL
F. SYNCHRO

FIGURE 9

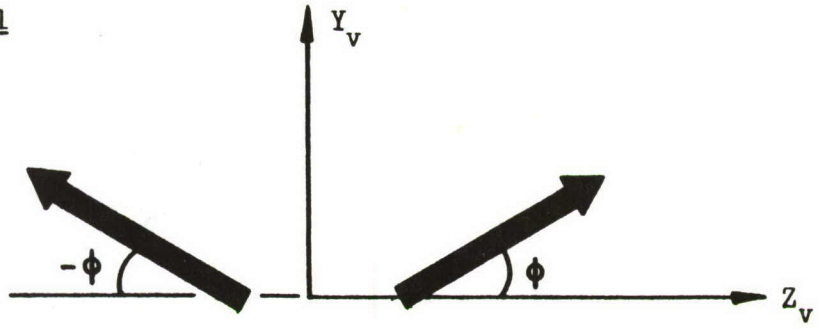
TWIN-GYRO CONTROLLER (SINGLE AXIS)

a. No Control Activation



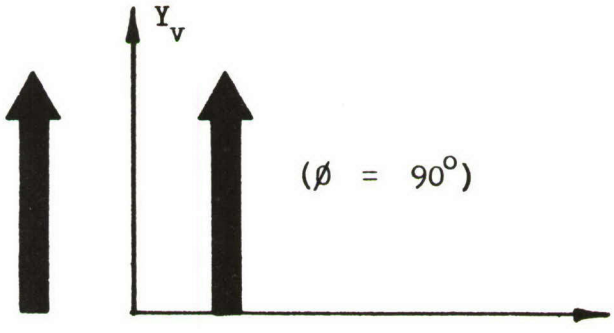
Y - momentum = 0
Z - momentum = $-H_1 + H_2 = 0$

b. Partial Control



Y - momentum = $-H_1 \sin(-\phi) + H_2 \sin \phi = 2H \sin \phi$
Z - momentum = $-H_1 \cos(-\phi) + H_2 \cos \phi = 0$

c. Control Saturated



Y = momentum = $H_1 + H_2 = 2H$
Z = momentum = 0

FIGURE 10

TWIN GYRO CONTROLLER PRINCIPLE

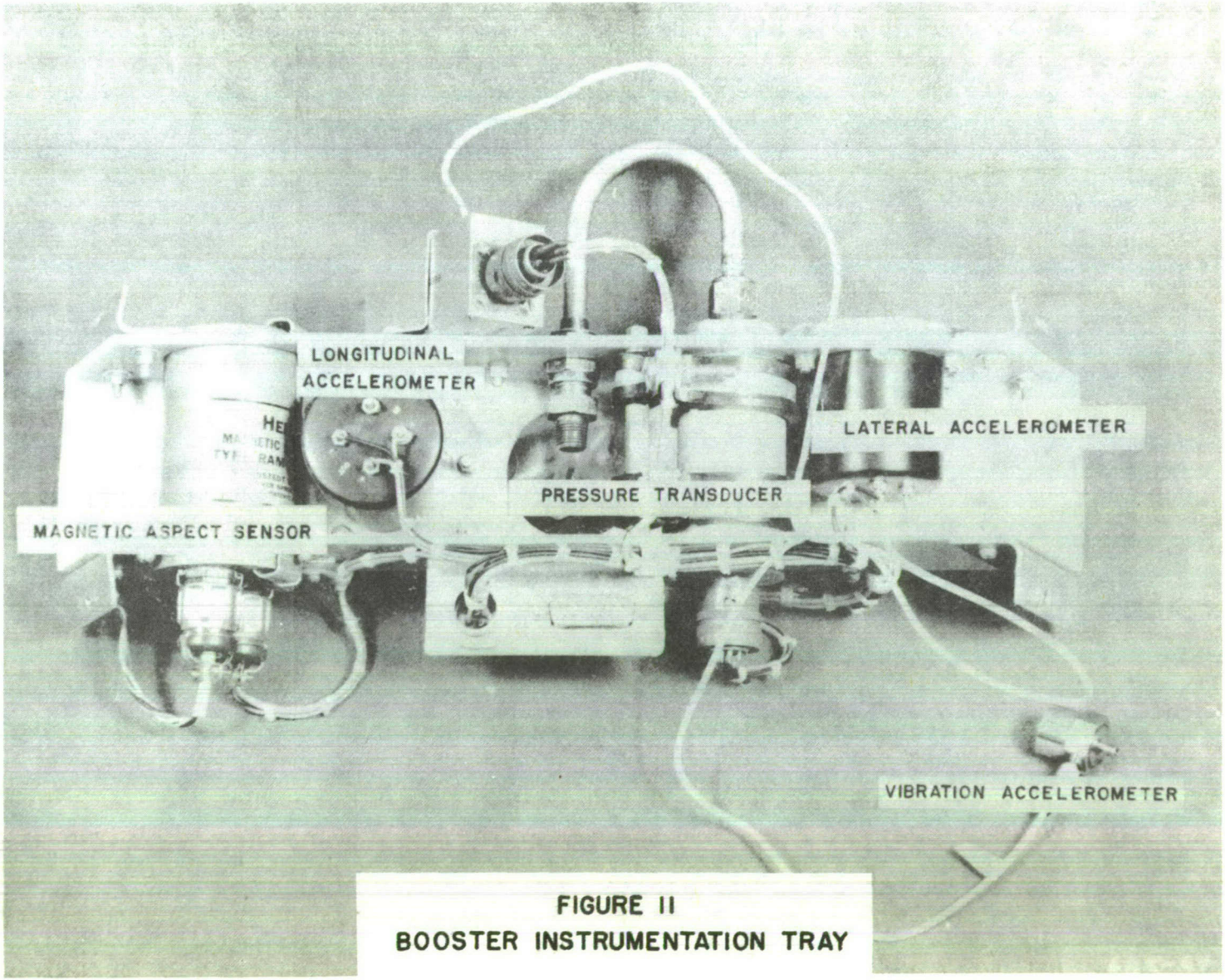


FIGURE II
BOOSTER INSTRUMENTATION TRAY

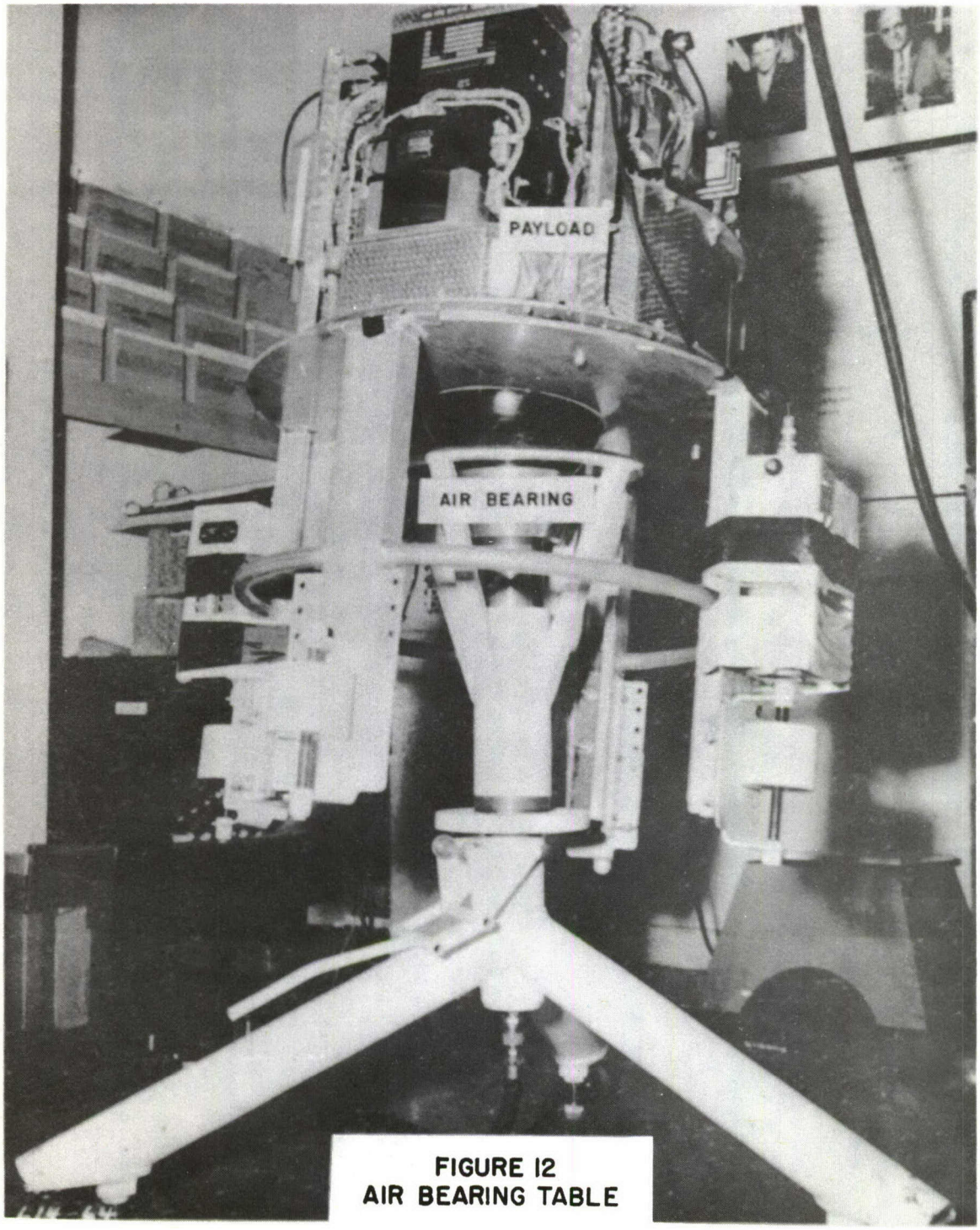


FIGURE 12
AIR BEARING TABLE

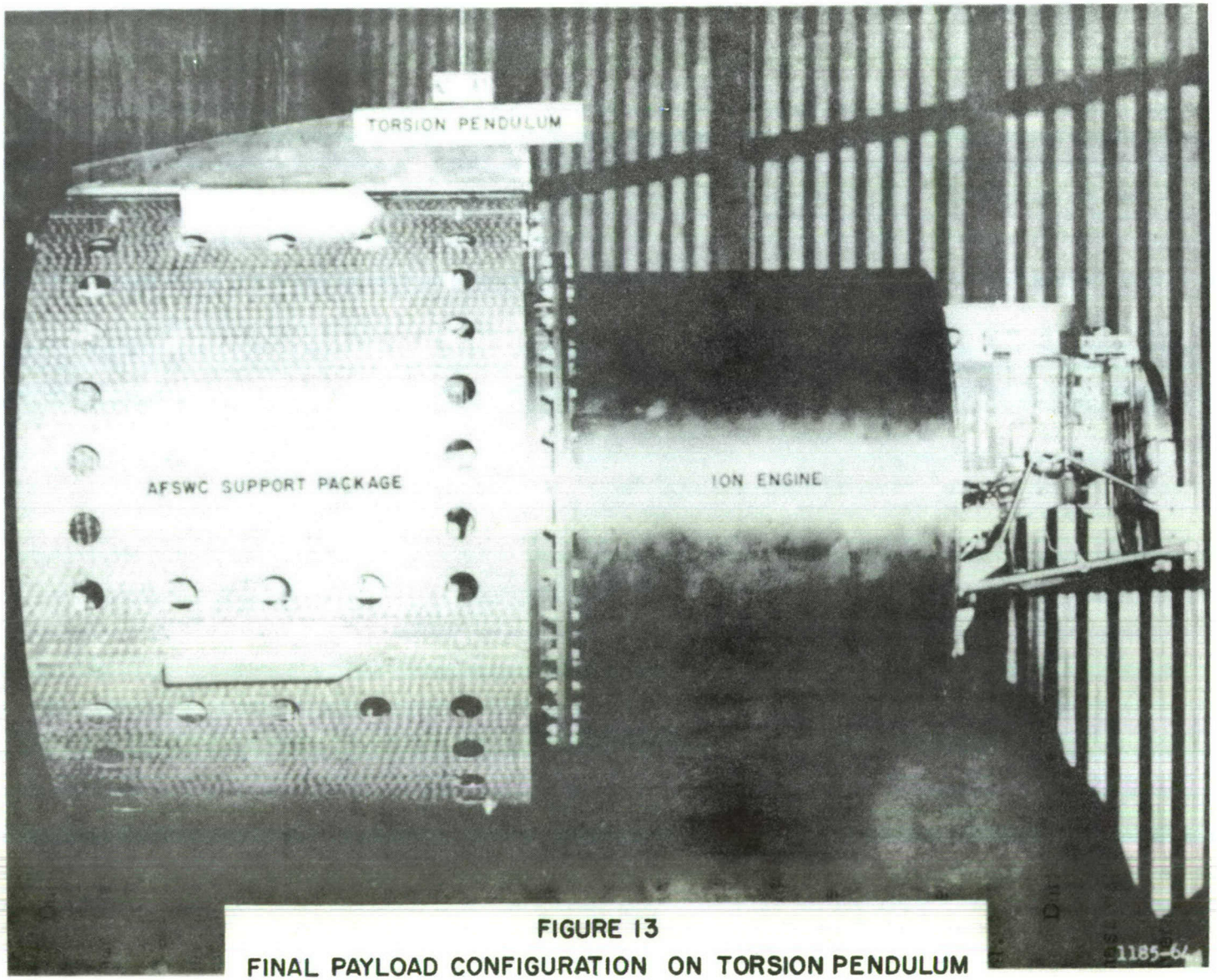


FIGURE 13

FINAL PAYLOAD CONFIGURATION ON TORSION PENDULUM

(U) CARDIOVASCULAR RESPONSES TO GRAVITATIONAL CHANGES AFTER PROLONGED BED REST

by

Major Perry B. Miller, USAF, MC
Lt Colonel Robert L. Johnson, USAF, MC
Lawrence E. Lamb, M. D.

Aerospace Medical Division
USAF School of Aerospace Medicine
Clinical Sciences Div, Internal Medicine Dept (SMKI)
Brooks Air Force Base, Texas

(U) CARDIOVASCULAR RESPONSES TO GRAVITATIONAL CHANGES AFTER PROLONGED BED REST

Major Perry B. Miller, USAF, MC, Lt Colonel Robert L. Johnson, USAF, MC, and Lawrence E. Lamb, M.D.

Prolonged space flight and prolonged bed rest are sufficiently analogous to justify concern that the tendency to faint upon standing observed in normal subjects after prolonged bed rest may also be present in crew members reexposed to the earth's gravity after a 30 day flight in the Manned Orbiting Laboratory. In both environments the need for physical activity is greatly diminished and significant positional changes in hydrostatic pressure in the body fluids are absent.

This paper presents a study of various effects of 4 weeks of absolute bed rest on the circulatory system of man. Its purpose is: 1. To provide data useful in the prediction of cardiovascular tolerance to gravity after landing from a 30 day mission in the Manned Orbiting Laboratory; 2. To provide a model for later use to compare the cardiovascular effects of 4 weeks of space flight to 4 weeks of bed rest; 3. To evaluate the efficiency of an antigravity garment in preventing postural intolerance after 4 weeks of bed rest, and by analogy, 4 weeks of space flight; 4. To determine the suitability of this period of bed rest for future studies seeking effective cardiovascular maintenance procedures adaptable to space flight.

Tolerance to headward (+Gz) and transverse (+Gx) acceleration before and after bed rest was also studied, but only a brief description of tolerance to a simulated Gemini reentry profile will be given in this paper.

METHODS

Subjects were 12 healthy male volunteers who had just completed 8 weeks of basic training in the United States Air Force. The same schedule of tests was followed before and after bed rest:

	A.M.	P.M.
Monday -	Blood Volume, Transverse G (+Gx)	
Tuesday -	Cardiac Catheterization - 6 subjects	
	Tilt Table - 6 subjects	
Wednesday -	Positive G (+Gz)	
Thursday -	Tilt Table - Antigravity Suit	Tilt Table - No Suit
Friday -	Tilt Table - No Suit	Tilt Table - Antigravity Suit
Saturday -	Tilt Table - Antigravity Suit	Tilt Table - No Suit
Sunday -	Tilt Table - No Suit	Tilt Table - Antigravity Suit

Blood volumes were determined before bed rest, after 11 days of bed rest, and after 4 weeks of bed rest by a dilution method employing radioactive iodinated serum albumin. In 6 subjects additional blood volumes were measured after 2 days of bed rest, and 15 and 29 days after the completion of absolute bed rest.

A transverse acceleration (+Gx) profile identical to that planned for the reentry profile in the Gemini flights was produced on the human centrifuge (Fig. 1). Peak acceleration was 10.6 G. The centrifuge was stopped if the subject failed to see a central light.

Cardiovascular responses to gravity were evaluated on a tilt table on which the subjects were suspended in a parachute harness at a 90 degree head-up tilt.

The subjects were tilted for 30 minutes or until syncope or marked presyncopal symptoms, such as loss of vision, occurred. Electrocardiograms and indirect blood pressures were recorded in the horizontal position before tilting and during each minute of the tilt. The antigravity garment was a modified partial pressure suit that provided 62 mm Hg. pressure to the lower extremities and the abdomen.

The results of the cardiac catheterization studies will be reported elsewhere.

During the 4 weeks of bed rest the subjects did not leave the bed, and sitting was prohibited. Baths were given by medical technicians. During the week of testing after bed rest the subjects were returned to bed after each test procedure.

Five tilt table tests were performed during the second week after bed rest, 2 during the third week, and 1 during the fourth week. In 6 subjects additional tilt table tests were obtained during the fifth and sixth week after bed rest.

RESULTS

Blood Volume. A loss in total blood volume, ranging from 200 ml to 1300 ml, occurred during 4 weeks of bed rest (Table I). The average loss for 12 subjects was 740 ml. Loss of plasma volume ranged from 300 ml to 800 ml with an average loss of 550 ml. Red cell mass decreased in 8 subjects, the decrease ranging from 100 ml to 500 ml. It was unchanged in 1 subject. Red cell mass increased 100 ml in 2 subjects and 200 ml in 1 subject. The average loss in red cell mass for all subjects during 4 weeks of bed rest was 180 ml.

In the 6 subjects in whom additional blood volume studies were performed, an average loss of 520 ml in plasma volume occurred after 2 days of bed rest. Loss

of plasma volume ranged from 300 ml to 600 ml. No significant change in red cell mass occurred.

In these subjects the average loss of plasma volume after 4 weeks of bed rest was essentially unchanged from that recorded after 2 days of bed rest. After 4 weeks of bed rest, red cell mass was decreased in 4 subjects, unchanged in 1 subject, and increased in another. The average loss in red cell mass was 160 ml.

During a 15 day period following completion of absolute bed rest, red cell mass decreased in all subjects, the loss ranging from 200 ml to 600 ml with an average loss of 370 ml. During this period plasma volume increased in 4 subjects, remained unchanged in 1 subject, and decreased in another.

Four weeks after completion of absolute bed rest 2 subjects still had significant reductions of total blood volume when compared to total blood volume before bed rest. Total blood volume was decreased 700 ml in 1 subject and 1000 ml in the other. Respective decreases in red cell mass were 500 ml and 800 ml.

Loss in total blood volume and loss of postural tolerance showed no quantitative correlation. For example, subject J, whose blood volume decreased only 200 ml during 4 weeks of bed rest, fainted 3 times out of four during tilting after bed rest and showed high orthostatic heart rates. In contrast, subject I, whose blood volume decreased 1300 cc, did not faint after bed rest and showed relatively low orthostatic heart rate.

Transverse Acceleration. During transverse acceleration no subject experienced

greater physical discomfort after bed rest than before bed rest. No change in ability to respond to a central light was observed. No significant cardiac arrhythmias were observed during or after acceleration.

Tilt Table Tests. Average resting heart rates were higher after bed rest than before bed rest in each subject studied (Fig. 2). Mean resting heart rate for all subjects was 58 beats per minute before bed rest and 71 beats per minute after bed rest.

The highest heart rate recorded during each head-up tilt was chosen as one index of orthostatic tolerance. Although this value varied rather widely from day to day in each individual tested, the range of highest orthostatic heart rates in subjects unprotected by an antigravity suit was clearly higher after bed rest than before bed rest (See Fig. 3 and Table II). The range of highest orthostatic heart rates recorded before bed rest bordered that recorded after bed rest in only one instance.

The range of highest orthostatic heart rates recorded during tilt table testing with an antigravity suit after bed rest was essentially no different from that recorded without an antigravity suit before bed rest (See Fig. 4 and Table II). Comparison of the range of highest orthostatic heart rates recorded during testing with an antigravity suit after bed rest to that recorded with an antigravity suit before bed rest shows that the antigravity suit does not eliminate completely the effects of bed rest on orthostatic tolerance (See Fig. 5 and Table II). Heart rates are higher after bed rest.

Twenty episodes of syncope occurred after bed rest in contrast to 7 episodes before bed rest (Table III). ~~One-third~~ 42% of the subjects, however, experienced no syncope after bed rest. In some tilt table tests blood pressure became unobtainable

at times, but the subjects did not faint. When these tests are added to those in which syncope occurred, an increased tendency for circulatory collapse after bed rest is more apparent (Table III). Two subjects experienced neither syncope nor a profound drop in blood pressure after bed rest.

When an antigravity suit was worn during tilt table testing after bed rest, syncope developed on one occasion only. During this test the subject complained of testicular discomfort from the parachute straps, a circumstance sufficient in itself to trigger syncope.

Six subjects fainted during tilt-table testing on the first day following resumption of physical activity on the ward (Table IV). After this, syncope on the tilt table was no more frequent than before bed rest. In the majority of the subjects for a number of weeks after bed rest orthostatic heart rates remained higher than those recorded before bed rest. (Table IV)

Lightheadedness was common during the first day of activity on the ward, but no subject fainted or experienced a severe presyncopal reaction. The subjects also complained of weakness, pain in the soles of their feet during walking, and soreness in the muscles of the legs and thighs.

DISCUSSION

The wide variation in loss of blood volume in different individuals during bed rest conforms with the reports of other investigators (1,4). A lack of correlation between loss of blood volume and loss of orthostatic tolerance was also observed by Taylor et al (4).

In the present study the greatest decrease in red cell mass was noted 8 days after resumption of full physical activity. Taylor et al. (4) reported an apparent

loss of red cells during the first week of reconditioning after 3 weeks of bed rest. In their study, plasma volume had returned to pre-bed rest levels after 1 week of reconditioning.

A major reason for the investigation of tolerance to transverse acceleration after prolonged bed rest was to determine whether dangerous cardiac arrhythmias would occur. The absence in the present study of adverse cardiovascular effects during transverse acceleration after bed rest adds for the first time a measure of objective support to the consensus of opinion that prolonged space flight will not decrease tolerance to the transverse acceleration of reentry.

The increase in resting heart rates after bed rest has been observed in other studies of bed rest (1,5). Sedentary individuals tend to have heart rates higher than those of trained athletes.

In previous studies of bed rest, the subjects were not continuously in bed. In one study weekly tilt table tests were performed, (1) and in others(3,5) bathroom privileges were allowed. The possibility that the brief exposures to gravity in the erect position might have counteracted to some degree the adverse effects of bed rest dictated the use of absolute bed rest in the present investigation.

In this study postural fainting on the tilt table occurred much more frequently after bed rest than before, but ~~one-third~~ 42% of the subjects did not faint. No subject experienced syncope during activity on the ward after bed rest, although occasional faintness was common. The effectiveness of contraction of the leg muscles in returning venous blood in the legs to the heart is well-known. Thus a prediction

based on this study of the astronaut's postural intolerance after a 30 day mission in space should not be overly pessimistic. A predisposition to fainting will probably be present, but the likelihood of fainting may not be as great as some believe.

Postural intolerance has been reported after American and Soviet space flights. It is not yet known whether a given period of space flight will produce a greater degree of postural intolerance than the same period of bed rest. Since postural intolerance has been demonstrated in only 4 days of inactivity at chair rest (2), one cannot conclude that the postural intolerance noted after space flights up to 5 days in duration augurs a more serious problem with longer flights than that observed during immobilization studies of similar duration on earth.

Previous studies showed the antigravity garment to be an effective countermeasure against the occurrence of postural syncope after 2 weeks of bed rest (3). The present study proves its equal value after the greater stress of 4 weeks of bed rest. The antigravity garment provides an important stopgap measure to protect the astronaut until the extent of the problem of prolonged weightlessness is known, and until effective inflight cardiovascular maintenance procedures are discovered.

A major objective of this investigation was to determine if 4 weeks of absolute bed rest would produce unequivocal evidence of a loss of postural tolerance in each person studied. One cannot assess the effects of space flight nor evaluate the efficacy of cardiovascular maintenance procedures without reliable criteria to judge changes in postural tolerance.

It is apparent from this study that postural tolerance both before and after bed rest varies from day to day in the individual. The problem of separating changes secondary to day to day variation in the individual from changes caused by bed rest

is solved by repeated tilt table testing. Susceptibility to syncope, before and after bed rest, varies greatly among individuals. In each subject, regardless of susceptibility to postural syncope, the distinct increase in orthostatic heart rates after bed rest provided a reliable criterium of loss of orthostatic tolerance.

SUMMARY AND CONCLUSIONS

1. Various effects of 4 weeks of absolute bed rest on the circulatory system were studied in 12 healthy male volunteers. An identical schedule of tests was followed before and after bed rest.
2. As judged by degree of physical discomfort, the retention of central vision, and electrocardiographic responses, tolerance to a simulated Gemini reentry profile on the human centrifuge was unaffected by bed rest.
3. An average loss of 740 ml. in total blood volume occurred after 4 weeks of bed rest, the loss ranging from 200 ml. to 1300 ml. An average decrease of 550 ml. in plasma volume and 180 ml. in red cell mass was noted.
4. Postural tolerance varied from day to day before and after bed rest. Repeated tilt table testing before and after bed rest allowed a more accurate appraisal of changes in postural tolerance caused by bed rest.
5. Postural syncope on the tilt table was more frequent after bed rest. ~~One-third~~ 42% of the subjects, however, did not experience syncope after bed rest. A comparison of the highest orthostatic heart rates recorded during each tilt table test before and after bed rest showed a distinctly higher range of orthostatic heart rates after bed rest in each subject studied.
6. When an antigravity suit was worn during tilt table testing after bed rest, postural syncope occurred on one occasion only. Orthostatic heart rates

tended to be no higher during testing with an antigravity suit after bed rest than during testing without an antigravity suit before bed rest.

7. These studies support the prediction that tolerance to transverse acceleration will be unaffected by a 30 day flight in the Manned Orbiting Laboratory. They suggest that postural tolerance after landing from prolonged space flight may vary widely among different individuals. They provide a model for later use to compare the cardiovascular effects of 4 weeks of space flight to 4 weeks of bed rest. They suggest that an antigravity garment should protect an astronaut from the development of postural syncope after landing from a prolonged space flight. They indicate that 4 weeks of absolute bed rest should prove an excellent part of any experimental design used for the evaluation of cardiovascular maintenance procedures adaptable to space flight.

REFERENCES

1. Deitrick, J. E., Whedon, G. D., and Shorr, E.: Effects of immobilization upon various metabolic and physiologic functions of normal men. Am. J. Med. 4:3, 1948.
2. Lamb, L. E., Johnson, R. L., and Stevens, P.M.: Cardiovascular deconditioning from immobilization at chair rest. To be published.
3. Miller, P.B., Hartman, B. O., Johnson, R. L., and Lamb, L. E.: Cardiovascular aspects of hypogravic states. To be published.
4. Taylor, H. L., Erickson, L., Henschel, A. and Keys, A.: The effect of bed rest on the blood volume of normal young men: Amer. J. Physiol., 144:227, 1945.
5. Taylor, H. L., Henschel, A., Bozek, J. and Keys, A.: Effects of bed rest on cardiovascular function and work performance. J. Appl. Physiol., 2:223, 1949.

CAPTIONS FOR FIGURES

Figure 1. Reentry profile for Gemini flights.

Figure 2. Average resting heart rates in horizontal position on tilt table for week of testing before and after bed rest.

Figure 3. Range of highest heart rates recorded during orthostasis in each tilt table test without antigravity suit performed during week of testing before and after bed rest. Lines within bars indicate heart rates within range. Duplication of heart rates is not shown.

Figure 4. Range of highest heart rates recorded during orthostasis in each tilt table test performed during week of testing before bed rest without antigravity suit and after bed rest with antigravity suit.

Figure 5. Range of highest heart rates recorded during orthostasis in each tilt table test performed with antigravity suit during week of testing before and after bed rest.

TABLE I

CHANGES IN CIRCULATING BLOOD

SUBJECT	*	Before	During Bed Rest			After Bed Rest	
		Bed Rest	2 days	11 days	28 days	15 days	29 days
A	TBV	4.9	4.2	4.7	4.5	4.7	4.8
	PV	2.8	2.2	2.4	2.4	2.8	2.8
	RCM	2.1	2.0	2.3	2.1	1.9	2.0
	HEMAT	43.1	47.5	48.3	47	41.5	41.9
B	TBV	6.0	5.0	5.5	5.4	4.9	6.1
	PV	3.4	2.8	2.8	2.9	2.9	3.6
	RCM	2.6	2.2	2.7	2.5	2.0	2.6
	HEMAT	43.1	44.7	48.9	46.7	40.2	41.9
C	TBV	5.5	5.4	5.1	5.3	4.8	6.1
	PV	3.2	2.9	2.7	2.8	2.9	3.5
	RCM	2.3	2.5	2.4	2.5	1.9	2.6
	HEMAT	41.2	46.1	46.3	45.9	39.7	42.1
D	TBV	6.0	5.4	4.9	5.4	5.1	5.3
	PV	3.4	2.9	2.6	3.0	3.1	3.1
	RCM	2.6	2.5	2.3	2.4	2.0	2.1
	HEMAT	43.1	47.0	46.6	44.2	39.1	40.0
E	TBV	5.6	5.2	5.2	4.5	4.4	4.6
	PV	3.2	2.7	2.8	2.6	2.8	3.0
	RCM	2.4	2.4	2.4	1.9	1.6	1.6
	HEMAT	42.6	47.2	46.2	41.7	36.1	38.2
F	TBV	4.8		4.4	4.3		
	PV	2.8		2.5	2.4		
	RCM	2.0		1.9	1.9		
	HEMAT	42.1		42.3	43.1		
G	TBV	6.3		5.9	5.7		
	PV	3.6		3.1	2.9		
	RCM	2.7		2.8	2.8		
	HEMAT	42.2		47.1	49.0		
H	TBV	6.4		5.8	5.2		
	PV	3.5		3.0	2.8		
	RCM	2.9		2.8	2.5		
	HEMAT	44.9		48.0	47.0		
I	TBV	5.5		4.7	4.2		
	PV	3.0		2.4	2.2		
	RCM	2.5		2.3	2.0		
	HEMAT	45.4		48.2	47.0		
J	TBV	4.5		4.6	4.3		
	PV	2.8		2.7	2.5		
	RCM	1.7		1.9	1.8		
	HEMAT	38.5		41.1	42.4		
K	TBV	5.8	5.3	4.8	4.7	4.6	6.0
	PV	3.5	2.9	2.7	2.8	2.9	3.6
	RCM	2.3	2.4	2.1	1.9	1.7	2.5
	HEMAT	40.2	45.1	44.2	39.9	36.7	40.8
L	TBV	5.5		4.9	4.5		
	PV	3.2		2.7	2.5		
	RCM	2.3		2.2	2.0		
	HEMAT	41.7		44.2	45.1		
MEAN ALL SUBJ	TBV	5.57		5.04	4.83		
	PV	3.20		2.70	2.65		
	RCM	2.37		2.34	2.19		
	HEMAT	42.34		45.95	44.92		
MEAN SUBJ A,B,C, D,E,K	TBV	5.63	5.08	5.03	4.97	4.75	5.48
	PV	3.25	2.73	2.67	2.75	2.90	3.27
	RCM	2.38	2.33	2.37	2.22	1.85	2.23
	HEMAT	42.22	46.27	46.75	44.23	38.88	40.82

*TBV - Total Blood Volume (Liters)
PV - Plasma Volume (Liters)
RCM - Red Cell Mass (Liters)
HEMAT - Hematocrit

TABLE II

HIGHEST ORTHOSTATIC HEART RATES
BEFORE AND AFTER BED REST

		Day of Testing					Day of Testing					Day of Testing					Day of Testing				
		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
WITHOUT ANTI-G SUIT	BEFORE	95	80	79	81	75	*	102	95	89	88	106	98	105	100	94	114+	110+	116+	119	117
	AFTER	125	130	116	120	127		150	158	168	155	133+	132+	137	142	130	155+	160+	141+	160+	147+
WITH ANTI-G SUIT	BEFORE		68	61	75	62		89	86	88	76		79	77	78	75		80	73	69	58
	AFTER		90	89	98	94		90	86	85	88		94	93	98	87		77	130+	95	102
	SUBJECT		A					B					C					D			
WITHOUT ANTI-G SUIT	BEFORE	*	105	88	91	80	*	101	84	86	79	*	98	90	92	127	122+	107	121	96	87
	AFTER		125+	149+	145+	148+		174	176	161	165		148	150	145	155	146+	147	152	135	154
WITH ANTI-G SUIT	BEFORE		68	56	65	66		70	63	66	65		60	68	71	74		83	81	86	76
	AFTER		90	95	90	96		100	84	103	107		120	80	96	112		108	102	101	92
	SUBJECT		E					F					G					H			
WITHOUT ANTI-G SUIT	BEFORE	85	83	87+	91	102	*	105+	97	89	112	*	107	150	88	84	112+	106	109	96	117
	AFTER	107	131	131	130	110		155+	164+	165	168+		150+	150+	154+	160+	159	172+	164	130	168
WITH ANTI-G SUIT	BEFORE		67	75	62	72		75	74	65	79		75	75	80	67		67	65	69	79
	AFTER		90	90	91	102		88	99	101	94		75	76	83	115		82	96	103	121
	SUBJECT		I					J					K					L			

*These subjects had cardiac catheterization on first day of tilt table testing.
+These subject experienced syncope during tilt table test.

TABLE III

SYNCOPE AND HYPOTENSIVE EPISODES
BEFORE AND AFTER BED REST

SUBJECT	NO SUIT										ANTIGRAVITY SUIT							
	BEFORE					AFTER					BEFORE				AFTER			
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	1	2	3	4
A																		
B																		
C						X	X											x
D	X	X	X			X	X	X	X	X	x				X			
E							X	X	X	X								
F										x								
G						x			x	x					x			
H	X					X	x	x							x			
I			X				x		x									
J		X				X	X	x	X									
K			x				X	X	X	X								
L	X					x	X											

X: Syncope

x: No blood pressure recordable - did not faint.

TABLE IV

ORTHOSTATIC TOLERANCE DURING RECOVERY FROM BED REST

	FIRST WEEK					SECOND WEEK		THIRD WEEK	FOURTH WEEK	FIFTH WEEK
SUBJECT	M	T	W	TH	F	M	F	F	TH	TH
A	109	114	122	115	107	106	100	103	103	104
B	125	116	119	114	122	136	118	108	104	108
C	107	117	114	121	107	100	105	103	112	118
D	134	125	167	142	150	132	154	128	125	139
E	141	115	142	128	112	135	117	110	135	134
F	139	125	*	92	104	90	95	68		
G	155	150	120	115	128	124	96	77		
H	165	118	106	126	112	105	109	109		
I	122	134	106	99	105	95	85	68		
J	178	142	114	117	97	129	141	104		
K	126	135	137	151	137	141	148	143	104	126
L	175	174	116	123	102	155	129	114		

Numbers are highest orthostatic heart rates during each tilt table test.
Underlined number indicatesoccurrence of syncope during test.

*Record Lost

GEMINI REENTRY PROFILE

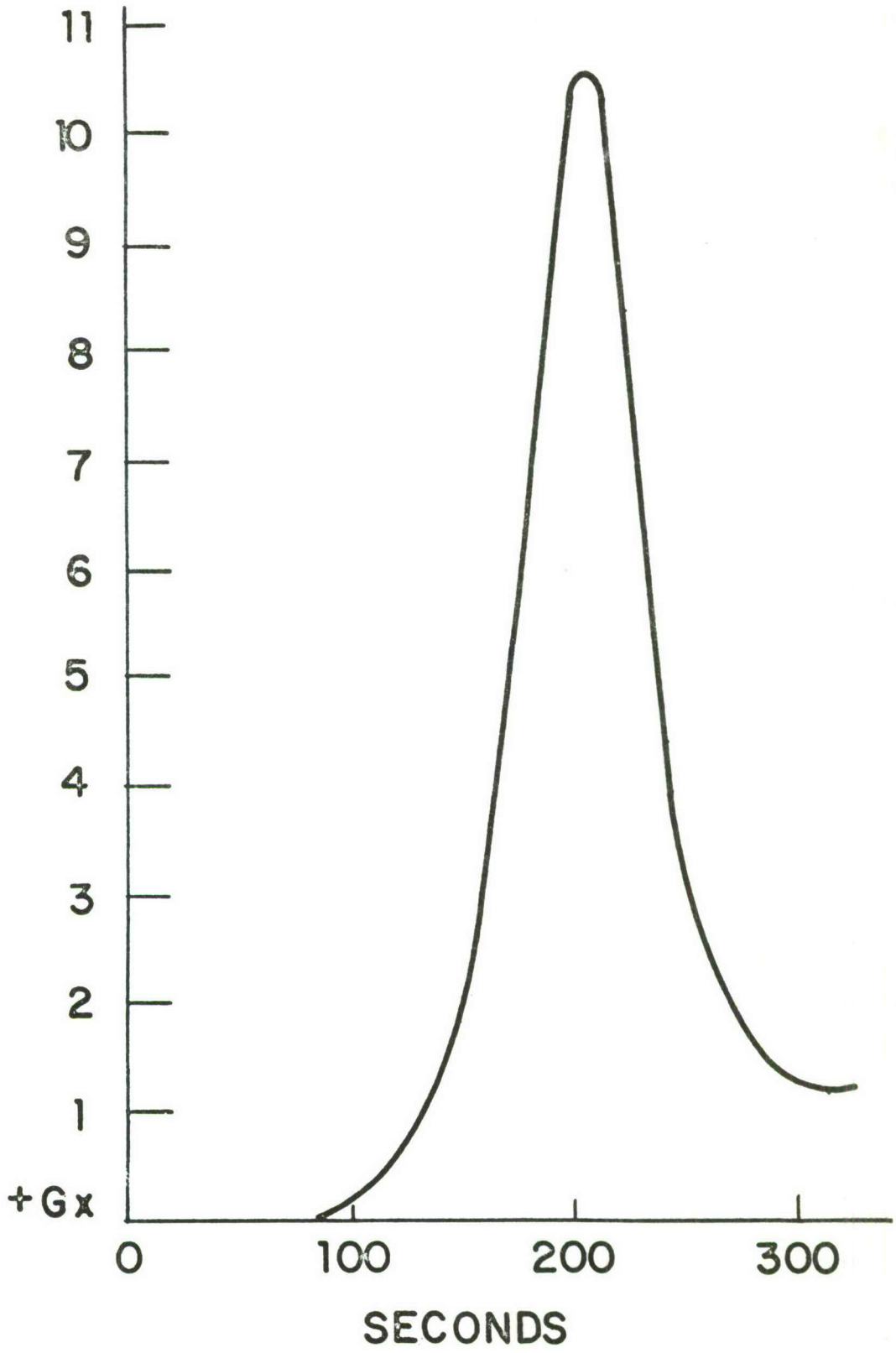
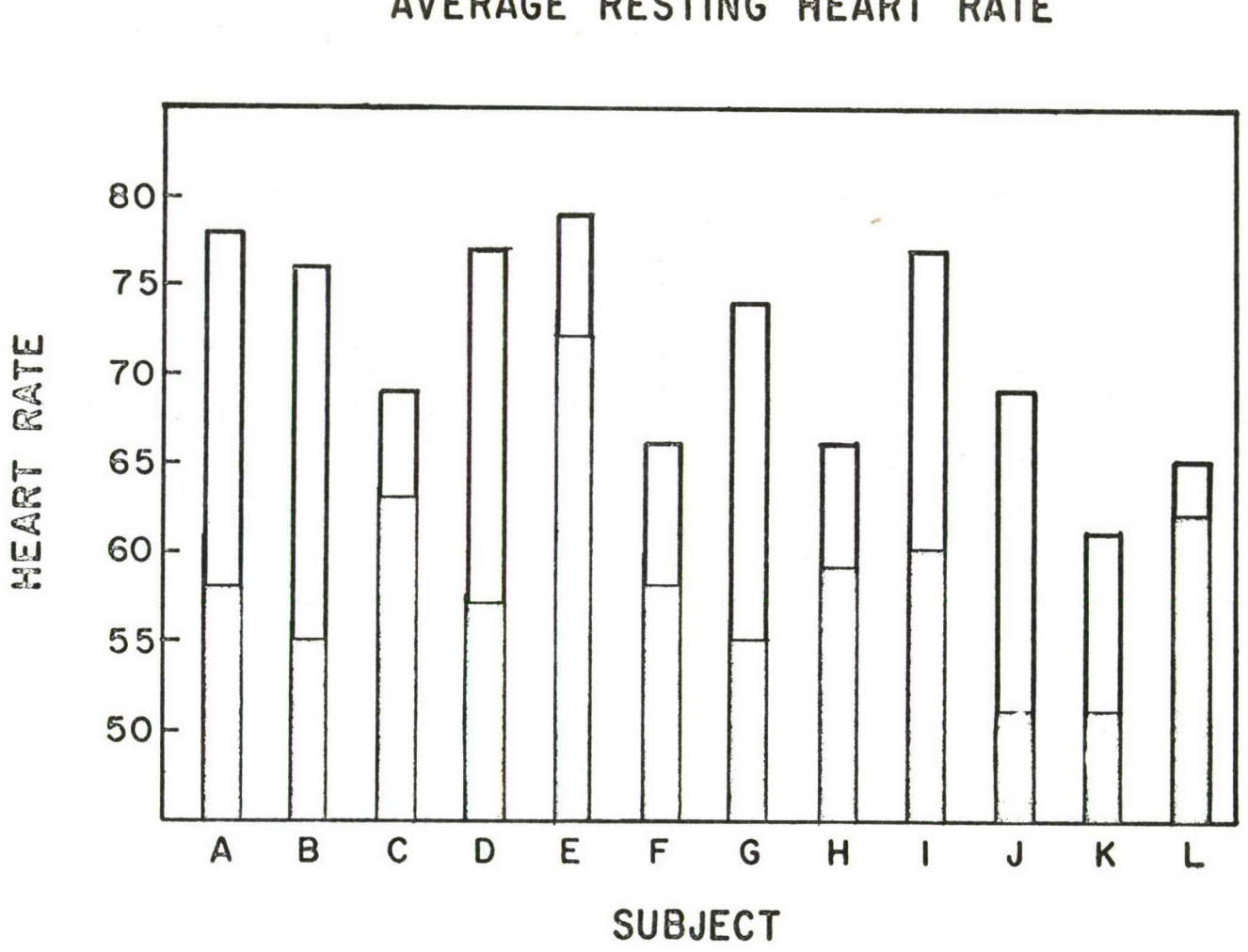


FIGURE 1



BLACK BAR: BEFORE BED REST

WHITE BAR: AFTER BED REST

FIGURE 2

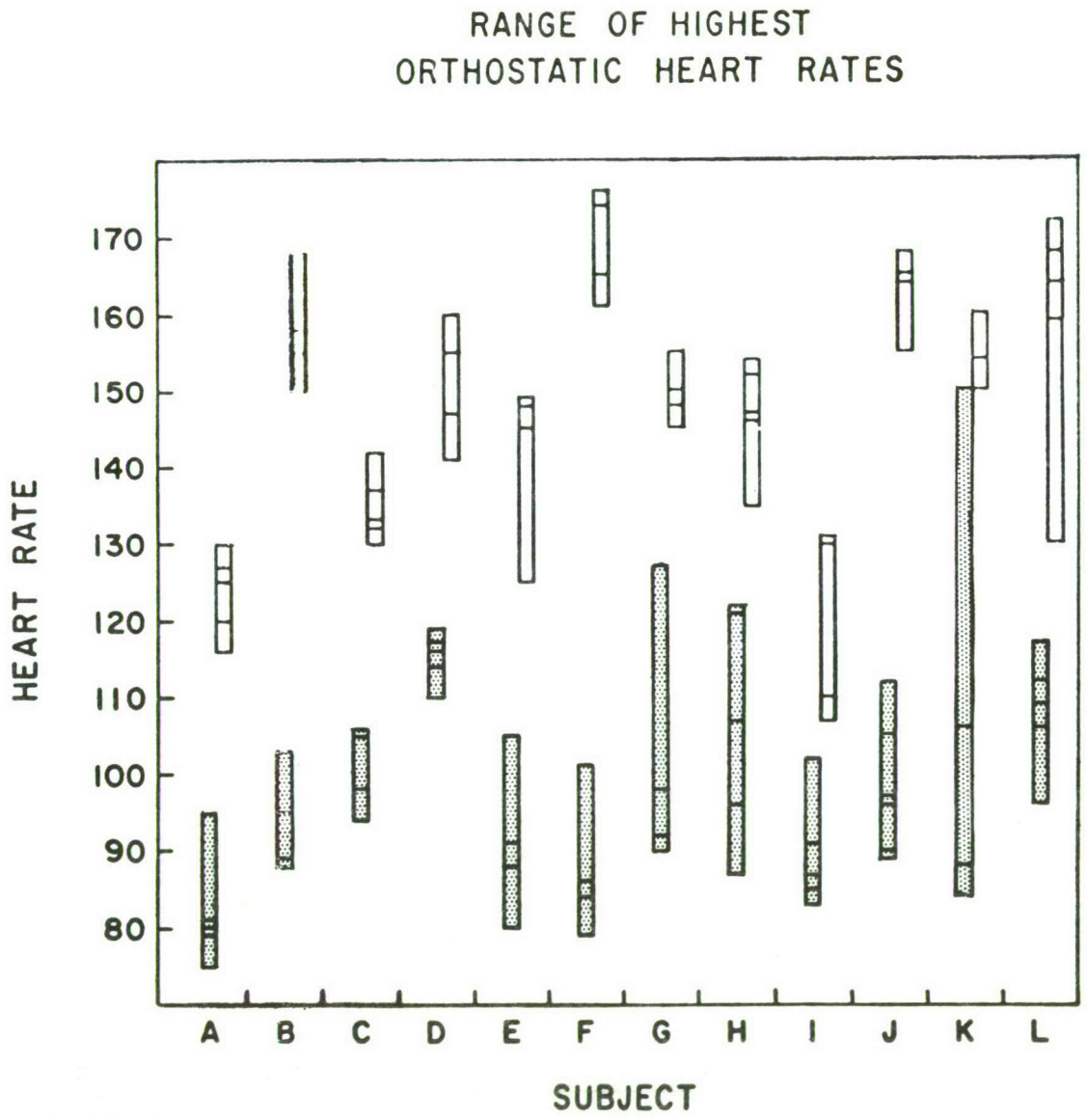
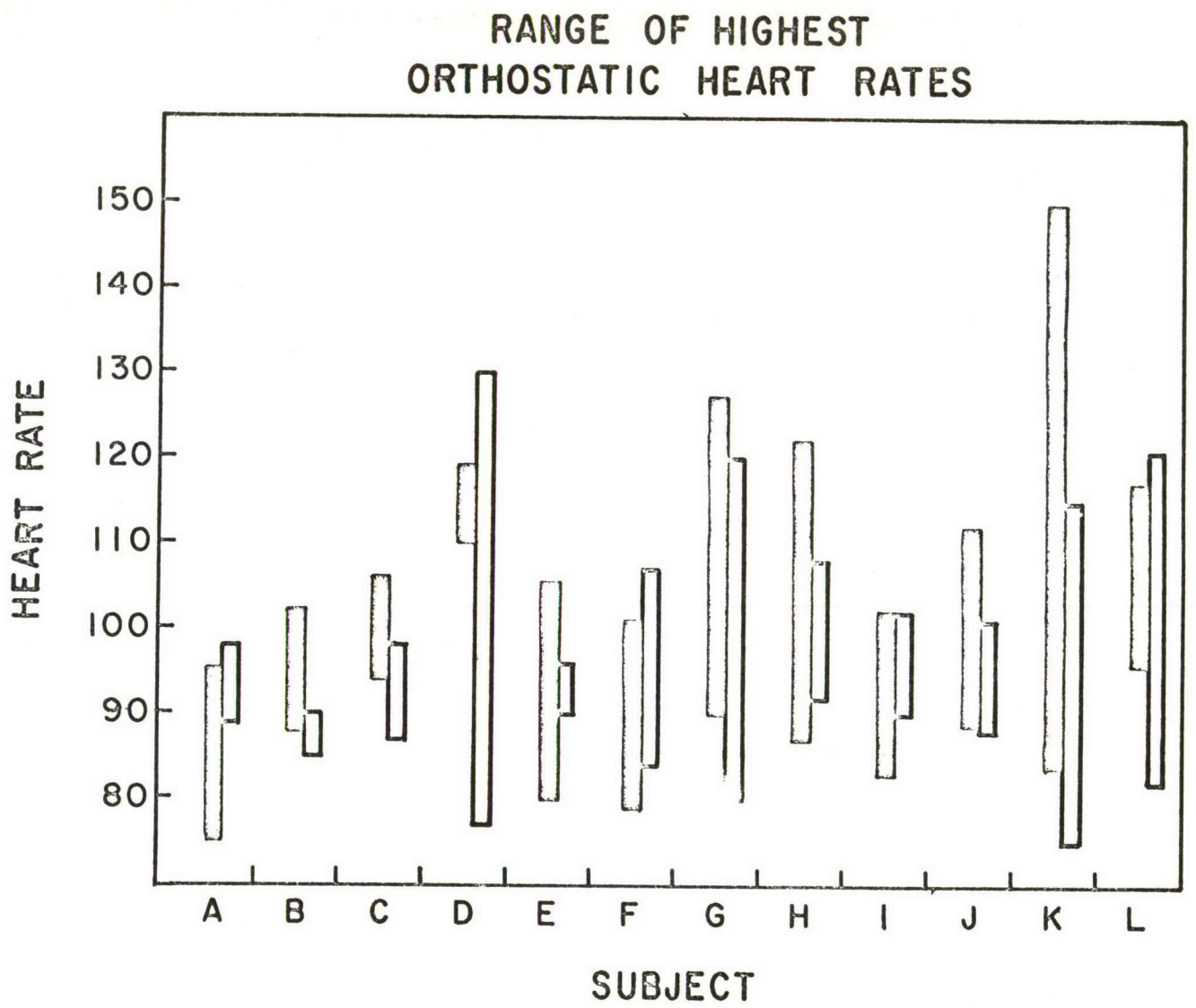


FIGURE 3 BLACK BARS: BEFORE BED REST – WITHOUT ANTIGRAVITY SUIT

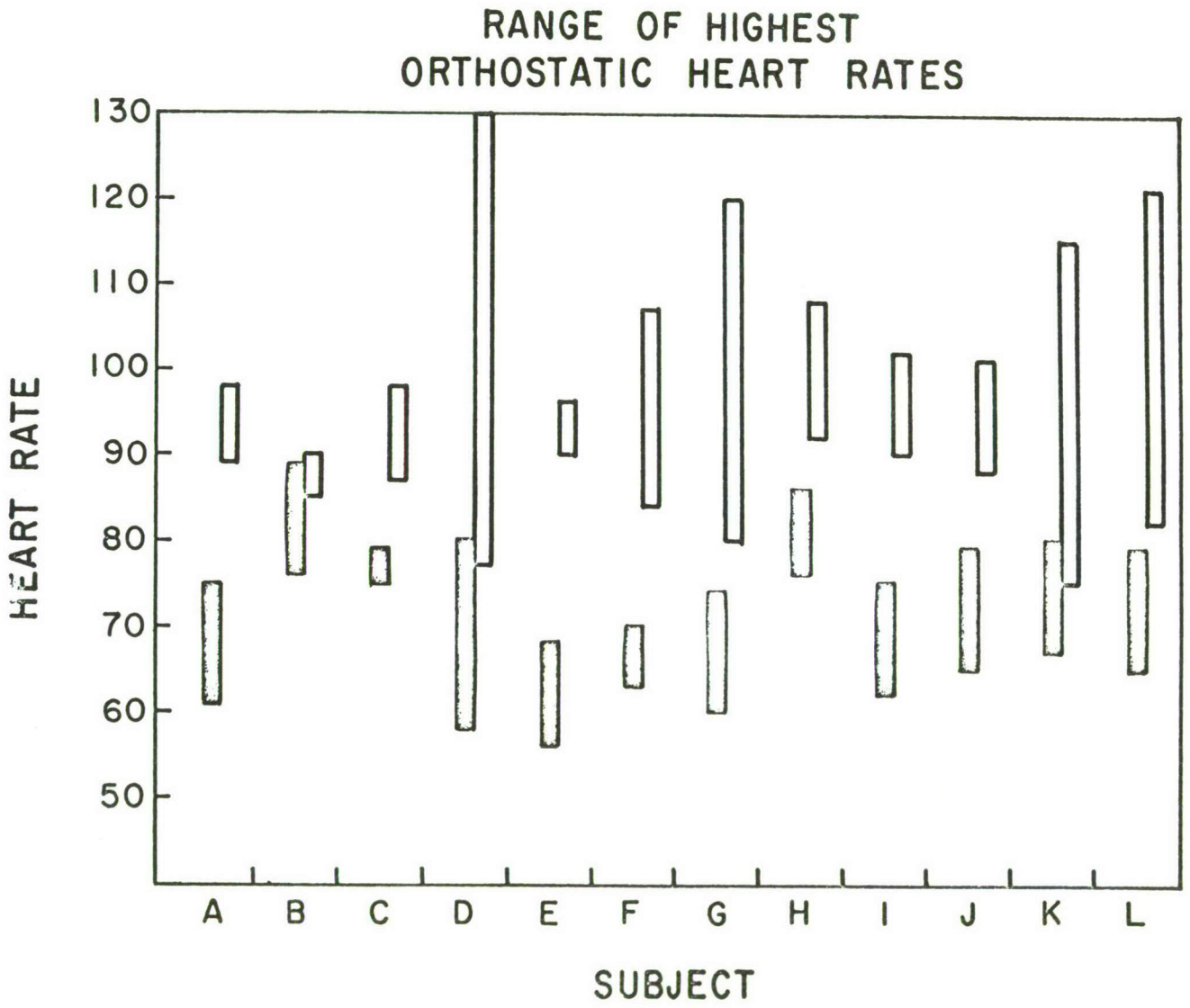
WHITE BARS: AFTER BED REST – WITHOUT ANTIGRAVITY SUIT



BLACK BARS: BEFORE BED REST – WITHOUT ANTIGRAVITY SUIT

WHITE BARS: AFTER BED REST – WITH ANTIGRAVITY SUIT

FIGURE 4



BLACK BARS: BEFORE BED REST – WITH ANTIGRAVITY SUIT

WHITE BARS: AFTER BED REST – WITH ANTIGRAVITY SUIT

FIGURE 5

INTERDISCIPLINARY MEASUREMENT OF HUMAN PERFORMANCE UNDER LOW AND ZERO GRAVITY CONDITIONS

DUANE FRANK KASTEN
GS-12

6570TH AEROSPACE MEDICAL RESEARCH LABORATORIES (MRPEC)
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

INTERDISCIPLINARY MEASUREMENT OF HUMAN PERFORMANCE UNDER LOW AND ZERO GRAVITY CONDITIONS

Legend has it that once upon a time a crew of three men guided their vessel out to sea on a strange voyage. Though the story of their journey has been lost in the mists of time, we presume that the interdisciplinary team of butcher, baker, and candlestick maker completed its journey successfully. With a mixed crew of experimental psychologists, anthropologists, physiologists, and engineers, the Aerospace Medical Research Laboratories at Wright-Patterson Air Force Base have sailed the sea of weightlessness since 1958. This team, too, is conducting its journey successfully.

Our journey into the realm of zero gravity began when two anthropologists suggested that one of their muscle strength studies could be greatly simplified if their test subjects could be made weightless. The anthropologists discussed their requirement for four seconds of weightlessness with a pilot-psychologist in the laboratory. He was intrigued with the idea and discussed zero gravity maneuvers with another laboratory pilot. When, very shortly thereafter, the zero G flights began this psychologist realized that human gross motion performance in weightlessness might well be of great interest to the Air Force in the near future. Gradually, over a two-year period, the psychologist transformed the Crew Stations Branch of the Human Engineering Division into an organization for measuring the capabilities and limitations of the weightless human.

The psychologists and anthropologists were soon collaborating on studies involving gross motions of the human body. We knew, for example, that a human floating in the aircraft free from all surfaces could not

propel himself without expelling mass unless he were influenced by an outside force, but, we asked, could he rotate his body to a new orientation without the help of external forces? To answer our question we enlisted the aid of a student engineer temporarily assigned to our Branch. After consideration of the zero gravity environment and the mechanics of the human body he declared that the human could indeed self-rotate or alter his orientation by means of limb manipulations. The analysis involved the application of theoretical mechanics to a rigid mathematical model of the body. Prediction of the expected degree of rotation was based on a quantitative evaluation of the model. As a result of his analysis, he proposed a set of motions that would permit a weightless man to change his orientation about three mutually perpendicular body axes. Armed with the battery of nine motions proposed by the engineer we turned again to our aircraft for an inflight validation of the self-rotation program. The zero gravity flights demonstrated the accuracy of the engineering analysis and proved that a weightless man did not have to depend completely on mechanical devices for controlling his orientation.

The brief example of the self-rotation program may be analyzed to show the methodology that has evolved in our zero gravity research. Generally a four part program develops for each major research area. This research program incorporates the following stages:

1. Problem Definition
2. Physical Analysis
3. Experimental Validation
4. Space Projection

Problem definition incorporates conception, formulation, and statement of the problem or task to be investigated. When we speculated about man's ability to perform self-rotation we were beginning the definition of a new problem. A zero gravity investigation may be undertaken because some previously unstudied phenomenon is noticed on the aircraft; or an investigation may originate as a logical outgrowth of previous research; or an investigation may result from armchair speculation as it did in the self-rotation study. The problem definition phase sometimes includes verification of the existence of the phenomenon or problem under weightless conditions. The final step in this first research phase is definition of the extent of the investigation to be undertaken and statement of the techniques of investigation.

Physical analysis is a step of extreme flexibility. The analysis may consist simply of a review of current knowledge of the effects of weightlessness on performance and application of that knowledge to the current problem; or the analysis may involve the application of specialized knowledge to a zero G problem as did our example of an aeronautical engineer producing pitch, roll, and yaw for the human vehicle; or the analysis may include a computer analysis of the problem and development of a mathematical model to explain the observed phenomenon. Regardless of the complexity of the analysis, its intent is to assist in the design of an experimental program for phase 3.

Experimental validation is generally the most lengthy phase of the research program. In this step the results of the physical analysis are subjected to test under weightless conditions in the zero-G aircraft.

Successful validation of the self-rotation analysis required 27 separate flights. This phase of the research program generally provides the bulk of the data which are transformed into design recommendation or performance capability reports.

Space projection involves extrapolation or adaptation of the findings of the experimental program to manned space operations. In the case of self-rotation, applications suggest themselves in the area of emergency maneuvers for astronauts outside their vehicles. Extrapolation to space applications may call for consideration of the effects of long term weightlessness on the data or for reinterpretation of the data considering the mechanics of an orbital situation.

Another example may serve to expand and clarify the importance of the space projection phase of our research. One aspect of human performance that interested us from the program's very beginning was man's ability to maneuver or locomote under weightless conditions. The necessity of moving about in the aircraft during zero-G flight would have been sufficient reason for giving thought to the problems of human locomotion. But we wanted, too, to consider problems that would arise eventually in larger space vehicles and problems that would confront an astronaut maneuvering on the outer surface of his vehicle or between two vehicles. Programs involving suction cup shoes, magnetic shoes, and adhesive shoes attempted to enable a man to walk while weightless. Programs designed to develop various types of personal propulsion units are continuing in our organization and have been seconded by extensive efforts in other organizations. But it was obvious that the fastest, surest, most economical means of

locomotion in the aircraft was soaring. With a minimum of experience, experimenters learned to propel themselves from a surface with the proper force exerted in the proper direction so that they simply floated to their destinations. Could this simple method of propulsion have applications in space? Perhaps, but a soaring astronaut should certainly be tethered to his vehicle by a strong, light line. So we investigated soaring and tether lines. We studied soaring speed. We studied soaring accuracy. We tested soaring techniques. We considered elastic tether lines and inelastic ones and shock absorbing lines. We looked at locomotion along a line back to a vehicle and retrieval of a disabled space worker. We were ready to send an astronaut from his capsule at the end of a long line.

A psychologist is a student of behavior; he is a measurer of performance; he is not an astronautical engineer. But when he learned that an astronaut attempting to return to his orbital vehicle along a tether-line could be forced into complex paths of motion that would intersect the vehicle only after long periods of time, or could, as he nears his vehicle, be subjected to velocities and accelerations that approach infinity, the psychologist realized again the value of the engineer to his research team.

An astronautical engineer who was added to our organization in 1960 performed the real work of projecting our findings into the environment of space. Without his efforts we would have been doing only half of our job. The psychologist can design experiments and measure human performance but often it is the engineer with his knowledge of orbital mechanics, coriolis effects, and mathematical analysis who makes our work valuable in a practical sense. In our example he demonstrated that " . . . the

magnitude of the retrieval problem depends primarily on the initial length of the tetherline and the initial tangential velocity of the [astronaut]. Reeling in a line a few hundred feet in length can result in forces which would be intolerable to an astronaut."

But there are many conditions in space that could be intolerable to an astronaut. Before 1958 many science fiction writers and even non-fiction science writers suggested that man could never tolerate space travel because, while weightless, he would have an endless sensation of falling. We know now, of course, that this doesn't happen. Some people do experience such a sensation very fleetingly as the aircraft reaches zero gravity but the feeling disappears immediately. It is akin to the momentary sensation one experiences when an elevator begins its descent. But even after this myth was laid to rest there was concern over an astronaut's ability to maintain his orientation under weightless conditions. Orientation is mediated primarily by vision, touch, and by that part of the inner ear known as the labyrinth. The otoliths of the labyrinth presumedly sense the direction of the gravity vector. If this were true, then a weightless man deprived of visual and touch cues would be completely disoriented. So we turned a twenty-five foot length of the aircraft into a darkened room and proceeded to test our hypothesis. Subjects, free floated in the aircraft away from cabin walls, floor, and ceiling, performed a series of motions or maneuvers while visual and touch cues were carefully controlled. The results supported the hypothesis. The weightless subjects could orient themselves by means of either visual or touch cues, but as soon as both were removed they literally didn't know which way was up. This experiment

provided evidence suggesting the loss of otolith function during weightlessness. But knowledge of the mechanism of the labyrinth is at best hazy. Once more we were treading upon ground at the very edge of the psychologists' knowledge. We needed the aid of a competent physiologist.

So the experimental aviation physiologist who joined our group started to systematize our knowledge of the effects of weightlessness on human physiology. Otolith function was one of his primary areas of interest. He informed us that, according to Captain Ashton Graybiel of the USN School of Aviation Medicine, one of the most sensitive indicators of otolith function is a phenomenon known as counterrolling. He explained that when the human head is rolled to the right or left, the subject's eyeballs rotate slightly in the opposite direction. Under one G conditions the eyes counterroll approximately 6 degrees when the head is tilted 90 degrees. Observation of this counterrolling phenomenon permits an appraisal of otolith function. Our physiologist conferred at length with personnel at the Navy School of Aviation Medicine in Pensacola, Florida. The Pensacola people are extremely interested in human orientation and they suggested that the counterrolling mechanism be observed under weightlessness. So we flew to Pensacola to continue our observations of orientation phenomena with Navy equipment and in cooperation with Navy personnel. And as anticipated, the counterrolling phenomenon disappeared at zero gravity.

By this time, of course, the physiologist was aware of the talents of the psychologists and engineers who were with him and after discussions

with a bioengineer he proposed an interesting extension of the counter-rolling study. He suggested that an engineer could devise a means for performing a dynamic analysis of counterrolling. Normally the subject's head is tilted at some selected angle and the eyes' counterrotation response is measured. In the proposed study the head would be tilted continuously in a specialized chair that rotated about a fore-and-aft axis passing between the subject's eyes. The counterrolling response would be measured continuously as the subject was rotated. Two engineers are now engaged in that effort. We anticipate that their techniques will furnish an improved method for evaluating the total otolith function. When the engineers have completed their investigation we will once more take the counterrolling study into the aircraft. This time we will utilize the new techniques to evaluate counterrolling as a function of gravity level. We hope to obtain a better estimate of the threshold of counterrolling. That is, the dynamic analysis should give us a better estimate of the lowest G level at which counterrolling appears. Successful completion of the program will not only add to our understanding of human performance under reduced gravity but will add to our knowledge of the basic physiology of the labyrinth.

The engineers engaged in developing this new technique are not assigned to our organization. They are graduate students at the Air Force Institute of Technology at Wright-Patterson. The counterrolling effort is their thesis research for a master's degree in electrical engineering. The Institute of Technology has been a most valuable source of ideas and personnel. The laboratories benefit by receiving the help of graduate

engineering students who serve temporarily as project engineers. The students benefit from exposure to the problems of practical research and they benefit by receiving advice and assistance in solving their thesis problems. During the past four years 14 students from the Institute have chosen reduced gravity problems as subjects for their master's theses.

One of the more interesting thesis projects currently under way illustrates once again the close working relationship between the engineer and the psychologists and anthropologists. As mentioned above, the psychologists quickly learned the value of a mathematical model of the body in their studies of gross motions, locomotions, and propulsion. Both the psychologists and our astronautical engineers wanted a model of the body that was more accurate and more generally useful than the one used in the self-rotation study. A graduate astronautics student undertook the task of developing such a model. First, twenty-five measurements were taken from each of 66 subjects. The measurements included body weight and body segment lengths and circumferences. A computer program was then written to construct an accurate fifteen segment model of each subject. Since the segments consisted of regular geometric shapes, an addition to the computer program permitted determination of the location of the center of mass of each segment and of the whole body. The moments and products of inertia were similarly determined for each model. The program computed these values for 31 different positions of the body and could do so for any body position chosen. These values were then compared with similar measurements made experimentally with the result that the computer program computed c.g. locations with a mean accuracy of $\pm$ 0.6

inches. Moment of inertia values were predicted with nearly equal accuracy. The computer program was then used to tabulate the center of gravity and inertia properties for Air Force men whose 25 measurements fall at the 5th, 25th, 50th, 75th, and 95th percentile of Air Force flying personnel. The results will be of primary importance in work with individual propulsion systems; cross-coupled rotations and fuel consumption will be minimized. But the computer program will be valuable in many planned studies involving gross motions of the human body.

The preceding example illustrates two additional aspects of our program that have not previously been mentioned. (1) Not all of our research lends itself directly to inclusion in the four-phase program described earlier and, (2) many problems of interest to the psychologist lend themselves to computer analysis. Another brief example will illustrate one instance of our using the analog computer facility at Wright-Patterson. Our early investigations of locomotion under zero gravity conditions included experiments with several rudimentary personal propulsion units. The earliest of these units consisted of a tank of compressed air, a length of high pressure hose, and a simple nozzle. Release of the compressed air was sufficient to provide three or four pounds of thrust. The thrust propelled the operator through the cabin but, since the nozzle was hand held, accuracy of thrust alignment with the man's c.g. was poor indeed. Experience improved the alignment accuracy, but the operator still had to achieve a proper pitch attitude before applying thrust. Before we pursued the program with more complex maneuvering units we wanted to try to compare the performance of subjects using pitch-thrust units with the performance of

subjects using units that provided thrust along orthogonal axes. Since we presumed that the maneuvering units would eventually be used in space, we wanted our comparison to consider the effects of orbital mechanics on the flight paths of our subjects. Obviously what we wanted was an orbital rendezvous simulation. And equally obviously, an analog computer would have to bear the burden of the simulation task. With the aid of the ASD Systems Dynamic Analysis Division such a simulation was programmed and run. We provided display and control information and designed the experiment and the engineers wrote the equations of motion and programmed the computer. We cooperated in training the subjects and in collecting the data. Though the qualitative aspects of the results were not unexpected, the magnitudes of the differences between the two control systems studied were larger than expected. The results indicated that in short range orbital rendezvous a self propulsion unit that provided thrust along orthogonal axes would probably be far superior to a unit that required the astronaut to select pitch attitude before applying thrust. Trained pilot subjects were able to rendezvous successfully on all trials with the former unit but on only 60 per cent of the trials with the latter control system. Not only was accuracy superior with the orthogonal axes system, but subjects achieved rendezvous with that system in shorter times than with the pitch-thrust system and used approximately the same amount of fuel. In addition, terminal deceleration was far more successful with the orthogonal control system. We like to think that this study had some influence on the design of subsequent, more complex maneuvering units. But perhaps transcending this possibility is the fact of understanding cooperation

from another discipline or talent in assisting us to fulfill our zero gravity research mission.

Throughout the history of our reduced gravity program, research personnel representing different disciplines have worked together in solving new and unusual problems. In 1958 there was no formal program for investigating the effects of weightlessness. Now, six years later, we have regular indoctrination flights, research under zero-G, Lunar G, even Mars G, pressure suited subjects practicing programmed space tasks and seeking the best methods for performing those tasks. We have research on human physiology, human performance, man-machine relationships, encompassing the cooperation and coordinated efforts of psychologists and anthropologists, physiologists, engineers, physicians and pilots: all members of the Air Force team dedicated to making this nation's space program the success that it must be.

(U) THE EFFECTS OF SECOBARBITAL AND D-AMPHETAMINE ON PERFORMANCE DURING A SIMULATED TACTICAL AIR MISSION

by

Richard E. McKenzie, Ph.D., Major, USAF, MSC
Lois L. Elliott, Ph.D.

Aerospace Medical Division
USAF School of Aerospace Medicine
Neuropsychiatry Dept, Psychobiology Branch (SMKN)
Brooks Air Force Base, Texas

(U)

THE EFFECTS OF SECOBARBITAL AND D-AMPHETAMINE ON PERFORMANCE DURING A SIMULATED TACTICAL AIR MISSION

Richard E. McKenzie, Ph.D., Major, USAF, MSC

Lois L. Elliott, Ph.D.

The operational deployment of Century Series fighter aircraft on extended overwater missions by the Tactical Air Command (TAC) poses some significant problems related to pilot proficiency under the effects of drugs. Specifically, the premission crew-conditioning program requires the pilot to be put to bed at an unusual hour in the late afternoon, to be awakened approximately eight hours later in the very early hours of the morning for his final briefing, and to be ready for takeoff shortly before sunup, about 3 hours later. The mission he flies lasts from 5 to 12 hours or more. During this time he has the usual demands made of him for precision pilotage, aerial refueling, and the like. He must perform his duties in relative isolation, with the usual discomforts of long hours in the cockpit of a demanding aircraft at altitude.

Early in the crew-conditioning program for these extended flights, it was found necessary to provide medication for inducing an adequate amount of restful sleep. This was not a universal requirement, since not all pilots requested medication. A TAC survey of some 200 pilots making these missions, however, showed that almost 70% requested sleep medication given in the form of 1 - 1/2 gr. of secobarbital, and three of the group requested a second capsule.

The pilots were also furnished two 5 mg. capsules of d-amphetamine (a stimulant) to be self administered in flight when they felt that fatigue might reduce their effective performance. Approximately 64% of the pilots surveyed took the stimulant along, and 24% actually used at least one capsule in flight after 4 hours. This time period corresponds to that reported by pilots as being the period of greatest fatigue. About the same number of pilots took a second d-amphetamine capsule roughly four hours after the first. Almost all (96%) of the pilots who took the stimulant reported that the results were satisfactory, while 83% who took the secobarbital for sleep reported that it had a desirable effect.

One operational use of medications to induce sleep and alertness in pilots performing an extended, demanding mission has been described. Within this framework a research study was designed to test the effects such medication might have on performance.

METHOD

Subjects

Our goal was to simulate the tactical air mission as closely as possible in terms of crew-conditioning, duration, and drug administration. An ideal situation would have been to use TAC pilots and Century Series simulators, but the cost in dollars and mission disruption made it impossible. Therefore, basic airmen selected in terms of general and mechanical aptitude (scores of at least 80 on the General Aptitude Index and 60 on the Mechanical Aptitude Index of the Airman Qualifying Examination)

were obtained from a pool awaiting further training at Lackland Air Force Base. Although these subjects may not have been equivalent to trained pilots, in many ways, they were the best substitutes available.

Apparatus

The task selected to simulate a pilotage operation, was the Multidimensional Pursuit Test (MPT) developed several years ago at the School of Aerospace Medicine as an aid to pilot selection. This task consists of a cockpitlike console with stick, rudder pedals, and a throttle quadrant control (fig. 1). Four instrument displays, simulating airspeed, turn, bank, and tachometer conditions are programmed to vary simultaneously, but independently, in direction and rate. The subject's task is to keep the indicators of these four instruments centered and "on target" by his manipulation of the proper controls. The measure of proficiency is the total time that all meters are simultaneously on target during a specified period.

The subjects were required to perform this task continuously for a 12-hour simulated mission. Brief 15-second rest periods followed each minute of operation. This, we felt, simulated the amount of time a pilot might be permitted to relax his attention and concentration on the task of flying.

Four MPT consoles were set up and carefully calibrated for score width and mechanical operation to minimize effects due to apparatus differences. An earlier pilot study, as well as a study by Payne and

Hauty (5), indicated that some sort of feedback which would indicate to the subject how well he was performing added to subject motivation. Therefore, the subject was provided with a counter which indicated when he was on target and gave a "score" for his own use. A light on the small counter panel indicated the scoring and 15-second "rest" periods.

Two MPT units were placed in one room and the other two in a similar room across a hallway. Both rooms had identical lighting fixtures and all windows were blacked out. Each operator's console was completely surrounded on four sides by white sheets suspended by hospital-type bedside stands. Subjects were not permitted to talk or communicate with each other in any way.

The subjects were required to wear standard Air Force headsets equipped with boom microphones to further isolate them. All communications between the experimenter's control station, located in an adjacent room, and the subjects were accomplished by means of this intercom system. If the subject wanted to contact "control," he pushed a button located on his control stick. This flashed a red light at the control station and the monitors could reply. All communications were recorded on magnetic tape.

Psychologic Testing

The following tests were administered to all subjects in an attempt to determine whether certain personality or aptitude factors, measured by these tests, might be related to performance factors or drug effects occurring in this experimental situation: (1) Cattell's Sixteen

Personality Factor Questionnaire (16PF); (2) six tests from Moran's Repetitive Psychological Measures (RPM); including Aiming (Form 8); Visualization (Form 10); Flexibility of Closure (Form 14); Perceptual Speed (Form 12a and b); Speed of Closure (Form 16); and Number Facility (Form 18).

In addition, a peer rating score from basic training and the AFQT (Armed Forces Qualifying Test) scores for each subject were obtained to evaluate as possible predictors of the MPT proficiency score.

Biochemical and Endocrine Measurements

Demand-void urine collections were made before, during, and after the simulated mission to evaluate the levels of urinary epinephrine, norepinephrine, and other execreted products indicative of possible physiologic changes. This aspect of the study was done in cooperation with Dr. H. B. Hale who, in a previous study (4), evaluated urinary products and reported a significant stress response in pilots flying 6-hour overwater missions in Century Series aircraft. This aspect of the present study is not yet complete and will be reported separately.

Physiologic Monitoring

Two subjects of each group were instrumented for ECG, heart rate, and respiratory rate measurements. Electrodes were of local manufacture, using silver - silver chloride material 1- 1/4 cm. in diameter and mounted in a plastic cup measuring approximately 3 cm. in diameter. The electrolyte used was EKG-Sol, manufactured by Beck-Lee Corporation. The

electrodes, filled with the electrolyte and attached by Eastman 910 adhesive, were positioned on the subject's chest for ECG and heart rate. Respiratory rate was determined by means of a thermistor placed in the airstream of one nostril. Leads from these sensors were fed into appropriate couplers of an Offner Type R Dynograph. Readout on heat-sensitive paper was continuous at 100 mm. per minute.

Drug Administration

The experiment was designed to approximate the operational situation in which (1) secobarbital is administered to induce sleep, and (2) d-amphetamine is taken to combat fatigue. Ingestion of d-amphetamine is preceded by about 9 hours of flying and pre-flight activity, while approximately 15 hours elapsed between ingestion of secobarbital and d-amphetamine. In the experimental laboratory it was necessary to investigate possible placebo effects as well as effects of receiving neither medication nor placebo. Therefore, six treatment groups were provided as indicated in table II.

Dose levels of 3 gr. of secobarbital and 5 mg. d-amphetamine were used. In the operational situation only a small percentage of pilots requested a second capsule of secobarbital, thereby raising their dose level from 1 - 1/2 gr. to 3 gr. It was reasoned however, that the larger hypnotic dose was available and that if secobarbital did not affect performance at the larger dose level, it would hardly be expected to do so at the more commonly used sedative dose of 1 - 1/2 gr.

(U)

Both of the drugs and the placebo preparation were obtained in No. 3 capsules and were identical in appearance. The prescribed drug for each subject was placed in a small pill box coded with the subject's number and by color for either nighttime or daytime administration.

Capsule administration was supervised by a technician who personally observed each subject to insure that the capsule was swallowed, and double blind procedures were followed.

Secobarbital was administered at 2100 hours (9 p.m.) of the second day, immediately prior to bedtime, while d-amphetamine was administered during the long test run of the third day at 1400 hours (2 p.m.). This schedule allowed 17 hours between drug administrations compared with an average for TAC pilots of 16 hours, permitting an adequate opportunity for performance decrement to occur.

Ten secobarbital and ten d-amphetamine capsules were selected randomly from the same manufacturer's lot and were evaluated by our Toxicology Department. This evaluation supported the presumed content designation, and amount.

While drug dosage was not varied according to individual body weight, it was important to be certain that no one treatment group was significantly different in weight from any other. The overall average subject weight was 159.93 lb. The average within-group difference from this mean weight was only $\pm$ 5.38. The greatest average difference was a +13.52 for the secobarbital-placebo group.

Experimental Design

The experiment was arranged in a basic 2 x 3 design with either seconal or placebo administered as the before-mission treatment and either d-amphetamine, placebo or control (nothing) administered as the during-mission treatment. Previous experience had indicated a need to control possible machine differences among the MPT consoles. Since four MPT devices were used, the 2 x 3 arrangement was enlarged to a 2 x 3 x 4 design yielding 24 experimental conditions. With a minimum of two replications, 48 subjects were required. Table III shows the randomized block design.

Subject Conditioning

Three days were required to brief and examine each subject group of 4 men. Day 1 was devoted to a detailed briefing as to the reason for the experiment and the nature of the "mission." While the subjects were not told which of them would receive drugs, or which drugs would be used, they were given a fairly complete briefing on the nature of TAC's operational problem and were told that we proposed to provide the answers through this experiment. The briefing was standardized and presented in a manner designed to elicit maximum support and motivation on the part of the subjects.

The need for a low-residue diet was explained in terms of the pilot's (and the subject's) inability to defecate while in flight. Other restrictions, such as no smoking and no beverages other than water, were

explained relative to the demand-void urine collection and analysis. The schedule, housing arrangements, and the fact that a technician would monitor them at all times were also explained. Two subjects were randomly assigned to a separate auditory test for a research study conducted by one of the authors (L.L.E.). The remainder of the day 1 was devoted to the administration of psychologic tests.

The morning of day 2 was devoted to a continuation of the auditory test on 2 of the subjects, while the other 2 received a 2-hour baseline test run for the physiologic monitoring. The afternoon was used to train all 4 subjects on the operation of the MPT apparatus. After a standardized practice session a 1-hour baseline MPT run was scored. In the evening, the subjects voided at 1700 hours (5 p.m.). At 2100 (9 p.m.), the first demand-void urine collection was made, drugs were administered, and the subjects were put to bed in separate rooms.

Day 3 was the day of the simulated mission. The "flight" began at 0700 hours and ended 12 hours later. The subjects were never told the length of their mission. They were told only that it would last over 6 hours. The reason for this was explained on the basis that a TAC pilot really did not know how long he would have to fly, since his destination could be weathered-in or changed in-flight. Actually, this was done so that the subjects could not attempt to pace their performance effort. Their watches were removed prior to "takeoff" to further enhance their sense of isolation, with the hope that subjective time would pass more rapidly and reduce the boredom of the operation.

(U)

The 2 subjects randomly selected for physiologic monitoring had electrodes attached prior to takeoff and were monitored continuously during the flight. The subjects' complete schedule is given in table I.

RESULTS

Proficiency Measurement

Since the MPT scores obtained in this study and an earlier pilot study revealed many differences in individual psychomotor aptitudes, a covariance analysis technic was used to control for MPT proficiency. The subject's scores during the last 30 minutes of the training period were considered indicative of his basic MPT performance and thus constituted the covariate. The MPT scores when corrected for the subject's end-of-training proficiency level will be referred to as "adjusted" scores.

MPT scores were recorded each 10 minutes during the entire 12-hour "flight." For analysis, 24 performance scores were obtained by averaging sets of three successive scoring periods. Therefore, each of the 24 scores represents mean performance during a half-hour period. These adjusted mean scores for each treatment group are shown in figure 2.

To further analyze the data, the 24 MPT scores were grouped arbitrarily into 4 periods. Period 1, the period of greatest performance decrement, included the first 4 hours. Period 2, the postlunch and leveling-off period, included the next 2- 1/2 hours. Period 3, the drug

administration period, included the next 1- 1/2 hours. Period 4, the amphetamine-effect period, included the remaining 4 hours of the run. These within-period results are shown in figure 3.

A further breakdown of the MPT performance results is shown in table IV. All scores have been pooled into 2 groups, secobarbital and placebo, and are shown at the beginning, the 4th hour, and the 7th hour. At the 12th hour, a breakdown of the average score values for the within-groups (d-amphetamine, placebo, and control) is also shown. The between-group MPT score averages were 650.6 for d-amphetamine, 440.9 for placebo, and 464.8 for control.

Analysis of the adjusted mean half-hour scores of figure 2, the within-period scores shown in figure 3, and the within and between group scores given in table IV shows essentially the same picture. We see that all treatment groups started at a relatively high level of proficiency, which degraded over 7- 1/2 hours. At this point no further decrements in performance occurred. The specific group either maintained this level of proficiency or improved.

The secobarbital groups (SP, SC, and SD) had consistently lower scores for the first 3 hours than the placebo groups (PP, PC, and PD). The group receiving d-amphetamine at 1400 hours (PD) showed a marked increase in proficiency starting at about the 15th scoring period (1430 hours, fig. 2 or period 3, fig. 3). The SD group did not start its increase as abruptly and, while improvement occurred, it did not reach the level of the PD group. However, while secobarbital may have

acted to increase the latency of the d-amphetamine response, the performance curve for the SD group might still be rising at the end of the "flight", while figure 2 shows that the d-amphetamine group had clearly "peaked" between 1730 and 1800 hours and underwent a remarkable decrement during the final hour. It is also interesting to note that the d-amphetamine group closely approaches, but does not reach, a proficiency level as high as that with which it started the test period. The effects of placebo and control treatments (SP, SC, and PP, PC) during the last 4 hours were not as obvious as one might desire. It is hard to speculate why the SC group shows improvement in performance during the last period of figure 3. Note that while the PP group also shows improved performance during this same period, figure 2 reveals that this group begins to degrade about mid-way during the period and then shows a considerable loss in proficiency toward the end of the run. Since the SC and PP groups tend to follow the same performance pattern, it seems fair to state that there was no apparent placebo effect and that the controls represented the middle ground between the enhanced performance related to amphetamine administration and the degraded performance maintained by those taking secobarbital and a placebo. One might speculate on the differential psychophysiological effects of lunch, but the present results do not justify any conclusions beyond those mentioned.

Figure 4 shows graphically the pooled mean MPT scores according to experimental treatment at the start and end of the 12-hour simulated

mission, and seems to illustrate, most effectively, the overall results obtained in this study.

The results of the complete covariance analysis are shown in table V. This is included to indicate the level of confidence with which one may regard the findings thus far discussed. As can be seen, the overall effect of secobarbital was significant at the .025 level. While the overall effect of d-amphetamine is reported as significant at only the .10 level, this is a statistical artifact produced by including the scores obtained before the drug was actually administered. A more direct test of the stimulant effect is the interaction between d-amphetamine and periods. This is significant at the .001 level. Figure 3 shows the increased proficiency produced during period 4.

While the overall difference between the MPT consoles (the machine effect) was not significant, there were two significant interactions - machine vs. time period 1 and machine vs. secobarbital. There appears to be no logical explanation of these significant interactions. One must be quite conservative, however, in interpreting an individual test of significance in an array of 48 such tests.

Physiologic monitoring

An examination of the ECG, heart rate, and respiration rate records of the 24 instrumented subjects reveals no apparent systematic changes related to the treatment conditions.

Physiologic testing

A multiple regression analysis was performed to assess possible relationships between MPT proficiency scores and psychologic test scores.

(U)

Prediction of subject proficiency based on the psychologic test scores was not significantly different from prediction based on the mean MPT proficiency scores alone. A second regression analysis was used to determine the possible relationships between these same psychologic test scores and change in MPT performance during the "flight." The difference in mean MPT performance scores for each subject between periods 1 and 2, representing the first four hours of flight (the time of greatest performance decrement), was used as a measure of performance change. The prediction of this decrement, using the same 13 test-score predictors and the binary variable of whether or not the subject had been given secobarbital, was not significantly better than prediction of decrement from the drug variable alone.

It is concluded that the characteristics measured by the 13 psychologic test scores used as predictors do not relate to proficiency factors required to operate the MPT device or to the occurrence of performance decrement.

Subjective Evaluation

Replies of 35 of the subjects to the standard debriefing questions were tabulated. Ten of the subjects who received secobarbital reported that they became sleepy during the "flight" and felt that their performance was poor, while only 6 of the placebo group made this same observation. Three subjects of the secobarbital group thought their bedtime medication had a "drug" effect, while 3 others of this group reported difficulty awakening in the morning. There were no such comments from the placebo group.

Five subjects in the secobarbital group reported they felt "uncoordinated" the morning of the "flight"; however, 3 members of the placebo group also reported feeling this way.

Of those receiving the d-amphetamine, 10 reported that they seemed to "wake up" or "were tired, but more alert" after taking the capsule, while only 3 of the placebo group made similar statements.

In spite of the relative isolation and the use of subjects unaccustomed to operator tasks of this duration, practically no reports of illusionary nature were elicited. One subject felt that the position of his console changed, but his was corrected by looking at the side screens; another reported that the dials on his panel "came together." Five others reported that they felt the equipment was not operating properly at times. None of these reports appeared to be related to the treatment group of the subjects.

The small number of subjective reports distributed over a rather wide variety of categories precludes accurate evaluation. Debriefing seemed to indicate, however, that drug effects which were demonstrated in performance changes could be, to some extent, subjectively evaluated by the subjects, particularly in the case of the stimulant.

CONCLUSIONS

This study reports two findings of interest relative to the effects of drugs on human performance. The first merely seems to confirm the enhancement of performance produced by d-amphetamine previously reported

by Payne and Hauty at the School of Aerospace Medicine and by others (1-3, 5-7). The second finding is perhaps of more importance since it demonstrates not only that individuals receiving secobarbital at bedtime show a performance decrement 10 hours later at the start of a simulated "flight," but also that even at the completion of their mission 12 hours later they continue to demonstrate degraded performance. Further, the administration of secobarbital appears to increase the latency and overall response characteristics of d-amphetamine when MPT performance is used as the criterion.

Even though the experiment was aimed at simulating at least some aspects of an actual flight mission, and the MPT device simulates some aspects of a pilot's job, it is not possible to generalize to the actual in-flight situation. This is true largely because there are no minimum standards of MPT performance that can be related to acceptable performance standards for aircraft. These restrictions, of course, apply to most laboratory test devices. To provide definitive data capable of answering the question - "Will this particular drug impair (or enhance) the pilot's ability to perform his required duties?" - it will be necessary to relate specific psychomotor or other laboratory test devices to specific pilot tasks. Theoretically, it should be possible to state with acceptable confidence that performance on laboratory task "X" at a certain score level indicates that pilotage under the experimental conditions being evaluated would be difficult, dangerous, or impossible.

(U)

Another possible solution to the problem of generalizing from the laboratory findings to the operational situation is the use of flight simulators. Even with flight simulators and qualified pilots to man them, however, the aeromedical scientist must still apply adequate criteria for identifying and reading out those aspects of performance related to safe, competent pilotage.

Until the necessary time, money, and scientific manpower are allocated and expended to yield results which can be validly generalized to the airborne human, we will necessarily have to be content with less and continue to be conservative in applying laboratory findings like these reported here to those who fly.

SUMMARY

This study was designed to investigate the effects of secobarbital and d-amphetamine upon performance during a simulated 12-hour "flight" mission. The results indicated that performance decrement occurred as a residual effect of secobarbital, using a psychomotor test as the criterion. This residual effect was present throughout the 12-hour test period which began 10 hours after the drug administration. Further, the residual effects of secobarbital apparently altered the response characteristics of d-amphetamine administered to one treatment group 17 hours after the original dose of secobarbital; however, the group receiving only d-amphetamine showed the enhancement of performance previously documented by other researchers.

(U)

The performance decrement obtained was not predictable by the psychologic test scores of the subjects, nor did it appear to be related to the gross physiologic measures obtained.

REFERENCES

1. Goodhart, R. S. (ed.) Modern drug encyclopedia and therapeutic index, 9th ed. New York: Ruben H. Donnelley Corp., 1936

2. Hauty, G. T., and R. B. Payne. Mitigation of work decrement. J. Exp. Psychol. 49:60-67 (1955).

3. Hauty, G. T., and R. B. Payne. Fatigue, confinement, and proficiency decrement. In Alpern, M., and M. Stern (eds.). Vistas in astronautics. New York: Pergamon Press, 1958.

4. Marchbanks, V. H., H. B. Hale, and V. P. Ellis. Stress responses of pilots flying 6-hour overwater missions in F-100 and F-104 aircraft. USAF School of Aerospace Medicine Report 62-112, Aug. 1962.

5. Payne, R. B., and G. T. Hauty. Effect of psychological feedback upon work decrement. J. Exp. Psychol. 50:343-351 (1955).

6. Payne, R. B., and E. W. Moore. The effects of some analeptic and depressant drugs upon tracking behavior. J. Pharmacol. Exp. Ther. 1115:480-484 (1955).

7. Payne, R. B., G. T. Hauty, and E. W. Moore. Restoration of tracking proficiency as a function of amount and delay of analeptic medication. J. Comp. Physiol. Psychol. 50:146-149 (1957).

TABLE I

(U) Subject's schedule

Day 1	Day 2	Day 3
0500	Reveille	Reveille; urine collection 0545 - Breakfast Group B electrodes attached
0600	0615 Breakfast	
0700 - Arrival at USAF SAM	Group A / Group B	Test run begins
0800 - Briefing	Audio test (separate study) / Baseline physiologic measurement	
0900 - RPM tests		Urine collection
1000 - 16 PF		
1100 - Lunch	Lunch	Lunch
1300 Group A Audio test (separate study) / Group B Unscheduled	Training on MPT	Urine collection
1400		Drug No. 2
1600 - Dinner	Dinner	Lunch
1700 - Unscheduled	Void; unscheduled	Urine collection
1900 - Unscheduled		Test run ends Debriefing
2000 - Unscheduled		Dinner
2100 - Bed	Urine collection; Drug No. 1; bed	Urine collection Return to Lackland AFB

TABLE II

(U) Treatment groups

Treatment		
Before mission	During mission	Symbol
Secobarbital	D-amphetamine	SD
Secobarbital	Placebo	SP
Secobarbital	Control	SC
Placebo	D-amphetamine	PD
Placebo	Placebo	PP
Placebo	Control	PC

TABLE III

(U) Randomized block design

SUBJECT	TREATMENT	MPT NO.	SUBJECT	TREATMENT	MPT NO.
3 (27)*		3 (2)*	6 (30)*		2 (3)*
9 (33)		1 (4)	12 (36)		4 (1)
16 (40)	SC	4 (1)	13 (37)	PC	1 (4)
22 (46)		2 (3)	23 (47)		3 (2)
5 (29)		1 (4)	2 (26)		2 (3)
11 (35)		3 (2)	8 (32)		4 (1)
18 (42)	SP	2 (3)	15 (35)	PP	3 (2)
20 (44)		4 (1)	17 (41)		1 (4)
1 (25)		1 (4)	4 (28)		4 (1)
7 (31)		3 (2)	10 (34)		2 (3)
14 (38)	SD	2 (3)	19 (23)	PD	3 (2)
24 (48)		4 (1)	21 (45)		1 (4)

* Numbers in parentheses indicate replication of design to obtain N = 8 in each cell.

TABLE IV

(U) MPT scores by period for major and within-treatment groups

Major group	Beginning (1st hour)	4th hour	7th hour	End (12th hour)	Within-group
Secobarbital	549.2	430.8	401.4	592.7	D-amphetamine
				347.2	Placebo
				494.3	Control
Placebo	664.9	483.1	456.9	708.5	D-amphetamine
				534.5	Placebo
				435.2	Control
S.D. ($\bar{d}$)*	39.9	40.6	41.8	69.1	
d.f.	44	47	53	44	

*Standard deviation of the difference between the means, with degrees of freedom.

TABLE V

(U) Analysis of covariance

Source	d.f.	M.Sq.	F	P
Secobarbital	1	1918383	6.09	-.025
D-amphetamine	2	964628	3.06	-.10
SxD	2	596625	1.89	NS
Machine	3	500969	1.59	NS
SxM	3	1566808	4.97	-.01
DxM	6	776673	2.47	-.10
SxDxM	6	176062	.56	NS
Regression	1	22880704	72.63	-.001
Error (A)	23	315034		
Period	3	1678181	42.71	-.001
SxP	3	62264	1.58	NS
DxP	6	277464	7.06	-.001
SxDxP	6	101131	2.57	-.025
MxP	9	14098	0.36	NS
SxMxP	9	25901	0.66	NS
DxMxP	18	39515	1.01	NS
SxDxMxP	18	21709	0.55	NS
Error (B)	72	39290		
Time/P1	7	371344	33.03	-.001
SxT/P1	7	15791	1.40	NS
DxT/P1	14	16281	1.45	NS
SxDxT/P1	14	5917	0.53	NS
MxT/P1	21	23863	2.12	-.005
SxMxT/P1	21	9973	0.89	NS
DxMxT/P1	42	4422	0.39	NS
SxDxMxT/P1	42	12476	1.11	NS
Error (C)	168	11241		
Time/P2	4	23583	3.62	-.01
SxT/P2	4	14393	2.21	-.10
DxT/P2	8	2022	0.31	NS
SxDxT/P2	8	11394	1.75	-.10
MxT/P2	12	5584	0.86	NS
SxMxT/P2	12	4564	0.70	NS
DxMxT/P2	24	8603	1.32	NS
SxDxMxT/P2	24	7994	1.23	NS
Error (D)	96	6523		
Time/P3	2	13491	2.97	-.10
SxT/P3	2	20051	4.41	-.025
DxT/P3	4	10232	2.25	-.10
SxDxT/P3	4	8205	1.80	NS
MxT/P3	6	6228	1.37	NS
SxMxT/P3	6	3200	0.70	NS
DxMxT/P3	12	4487	0.99	NS
SxDxMxT/P3	12	7004	1.54	-.10
Error (E)	48	4549		
Time/P4	7	41189	4.67	-.001
SxT/P4	7	20755	2.35	-.05
DxT/P4	14	9047	1.03	NS
SxDxT/P4	14	7813	0.89	NS
MxT/P4	21	7259	0.82	NS
SxMxT/P4	21	10627	1.20	NS
DxMxT/P4	42	6114	0.69	NS
SxDxMxT/P4	42	10371	1.17	NS
Error (F)	168	8826		

Minus sign (-) in front of probability figure indicates "less than." Time periods P1 through P4 are referred to in text.

LEGENDS

FIGURE 1

(U) Picture of subject at MPT console.

FIGURE 2

(U) Graph of adjusted MPT scores of treatment groups by scoring periods and time.

FIGURE 3

(U) Graph of MPT scores of treatment groups showing adjusted mean scores for the selected analysis periods.

FIGURE 4

(U) Pooled mean MPT scores according to treatment at the start (S) and end (E) of the 12-hour mission.

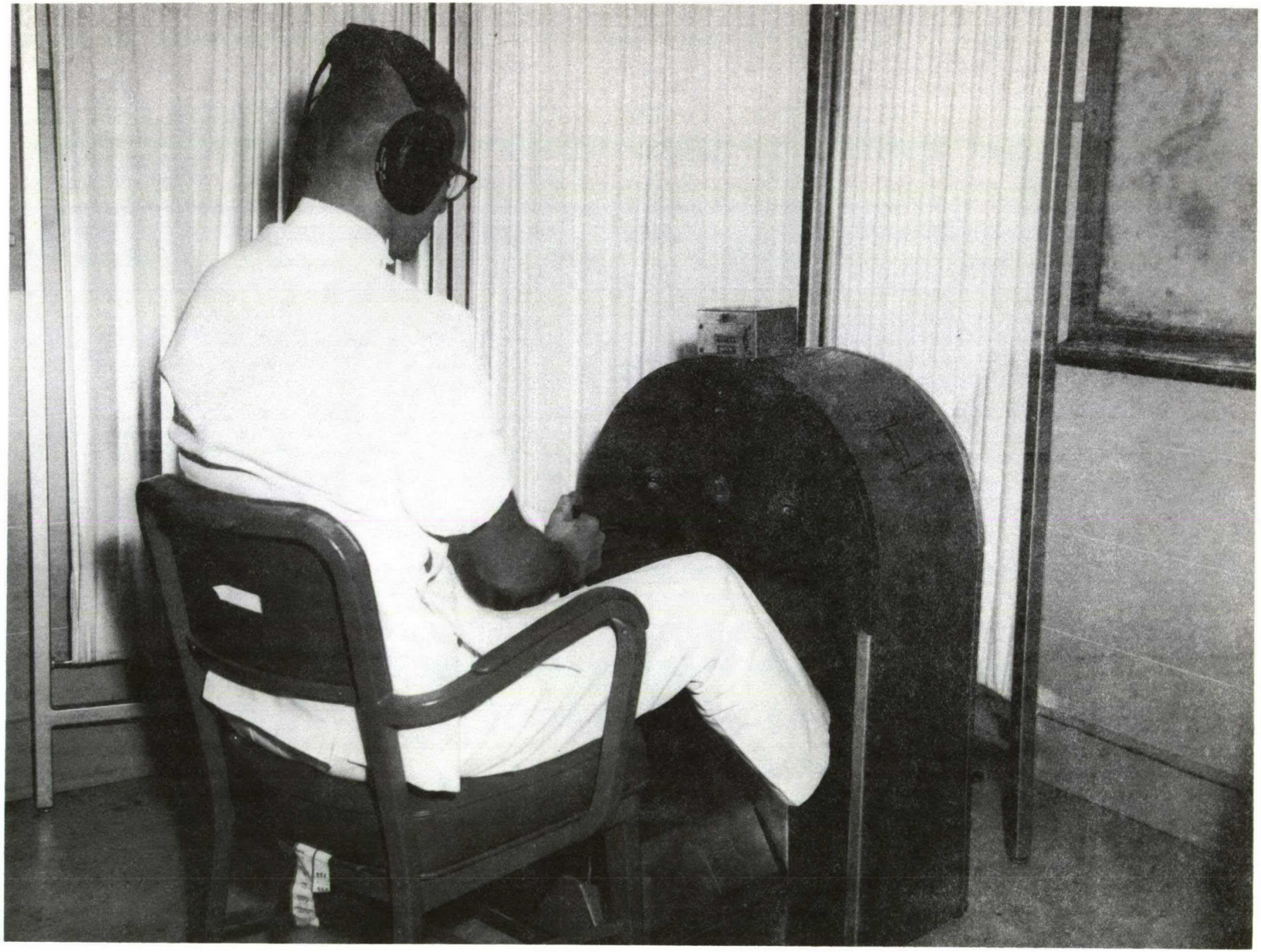


FIGURE 1

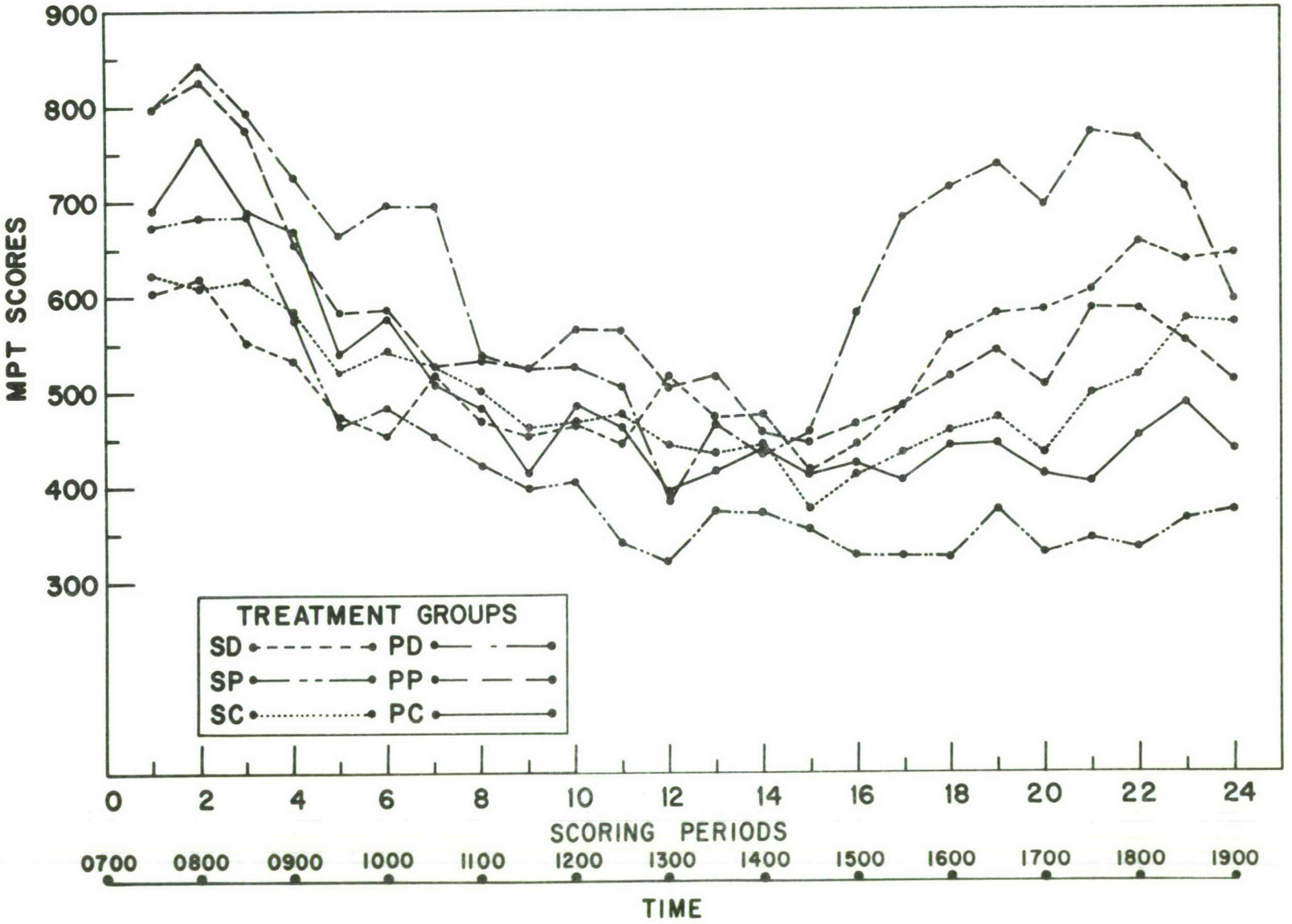

MPT SCORES
900
800
700
600
500
400
300
TREATMENT GROUPS
SD
PD
SP
PP
SC
PC
0
2
4
6
8
10
12
14
16
18
20
22
24
SCORING PERIODS
0700
0800
0900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
TIME

FIGURE 2

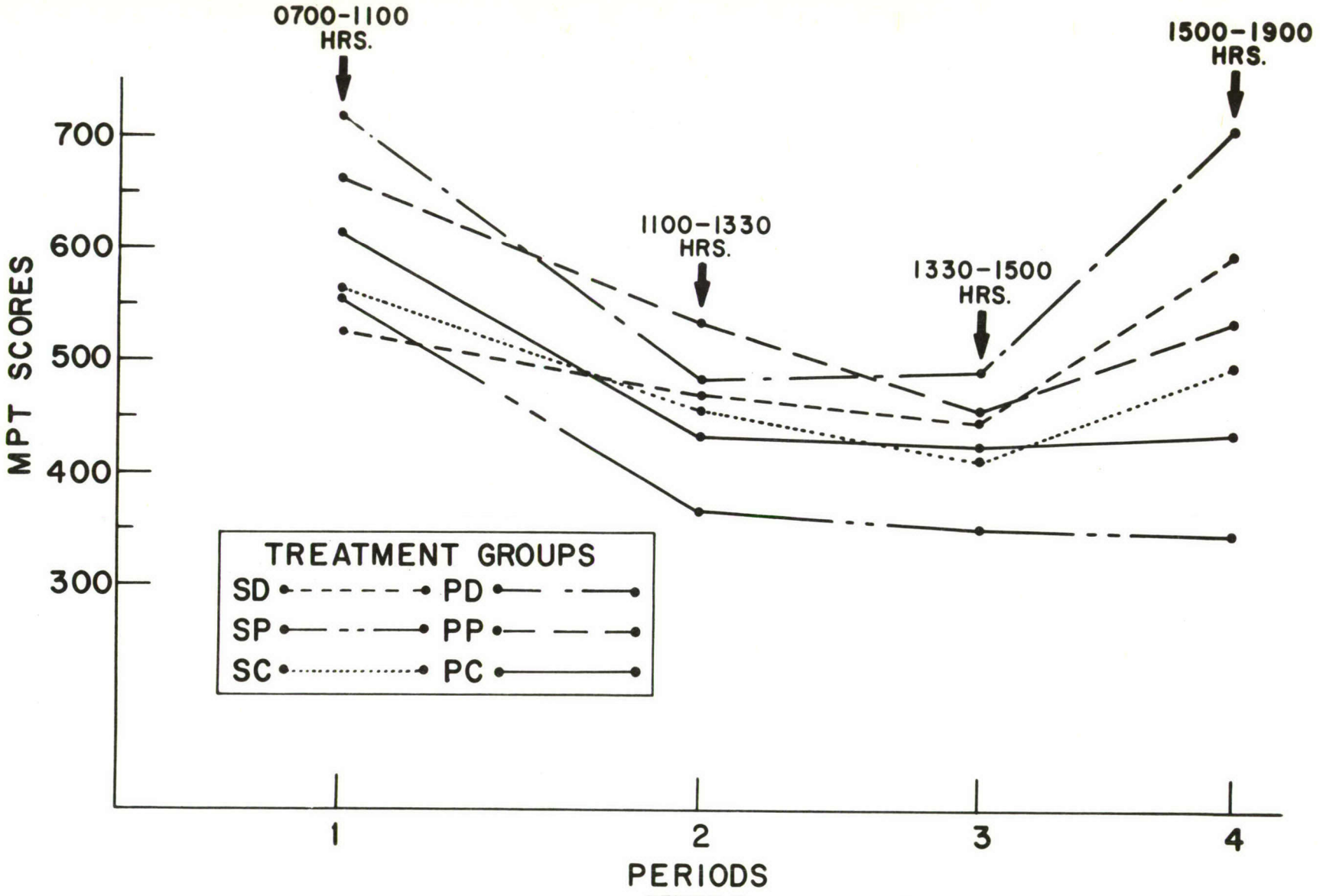

0700-1100
HRS.
1100-1330
HRS.
1330-1500
HRS.
1500-1900
HRS.
MPT SCORES
700
600
500
400
300
TREATMENT GROUPS
SD
PD
SP
PP
SC
PC
1
2
3
4
PERIODS

FIGURE 3

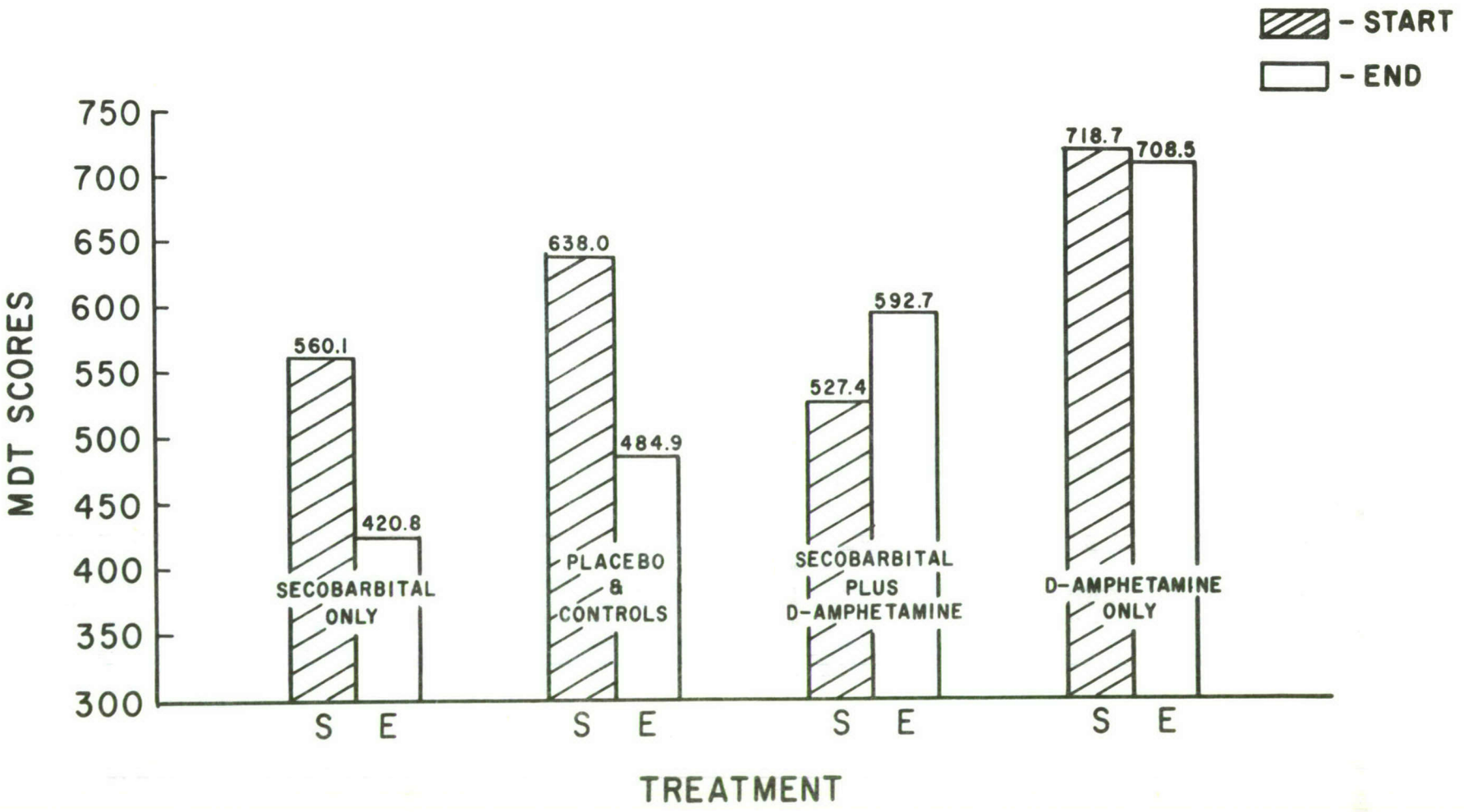

- START
- END
MDT SCORES
750
700
650
600
550
500
450
400
350
300
560.1
420.8
638.0
484.9
527.4
592.7
718.7
708.5
SECOBARBITAL ONLY
PLACEBO & CONTROLS
SECOBARBITAL PLUS D-AMPHETAMINE
D-AMPHETAMINE ONLY
S E
S E
S E
S E
TREATMENT

FIGURE 4

DIRECT ELECTROFLUIDDYNAMIC ENERGY CONVERSION PROCESSES FOR POWER GENERATION

by

M. O. Lawson
H. J. P. von Ohain

Aerospace Research Laboratories
Office of Aerospace Research
United States Air Force

ACKNOWLEDGEMENT

The authors wish to acknowledge the many conceptual and analytical contributions made by Dr. F. Wattendorf*, who for many years collaborated with the authors on the electrofluiddynamic inhouse research program.

*Vice Chairman, Advisory Group for Aerospace Research and Development (AGARD-NATO), Paris

I. INTRODUCTION

This paper is devoted to electrofluiddynamic (EFD) energy conversion processes. In these processes, fluid dynamic energy is directly transformed into electrical energy by passing a working medium containing ions or charged colloids through an electrostatic field. This contrasts with magnetofluiddynamic (MFD) energy conversion processes, whereby an electrically conductive working medium is passed through a magnetic field.

EFD processes are not to be considered as competing with MFD processes, but rather as complementing MFD processes in the energy conversion spectrum. In general, MFD processes must employ very high gas temperatures to obtain a sufficiently high electrical conductivity, and are characterized by high current density and a relatively low voltage; EFD processes, on the other hand, can employ low gas temperatures and are characterized by low current density and very high voltage.

The basic idea of employing EFD processes for power generation is not new; investigations on several aspects of EFD energy conversion processes have been reported (Refs. 1, 3, 4, and 6). The purpose of this paper is to treat comprehensively the entire energy conversion cycle from heat into electrical energy. In the first part of this paper, the EFD performance and scaling characteristics are given, whereby the effects of aerodynamic drag, slip, physical properties of the working gas and geometric parameters are taken into account. In the second part, aerothermodynamic processes are treated which overcome the inherent incompatibility between the electrofluiddynamic and thermodynamic cycle requirements.

II. QUALITATIVE DESCRIPTION OF THE EFD ENERGY CONVERSION PROCESS

The basic EFD process, which is considered in this paper, can be described briefly in a qualitative manner by referring to Fig. 1: electrically charged particles of one polarity (indicated by dots) are transported by fluid dynamic energy from the grounded entrance electrode into the collector electrode against the electrostatic field in the conversion section. This process corresponds to the functions of the mechanically driven belt in a Van de Graaff generator. However, the gas speed in an EFD process can be far higher than the belt speed of a Van de Graaff generator; also, the electrically charged particles in an EFD process are distributed throughout the volume flow of the working medium rather than being concentrated on the belt surface of a Van de Graaff generator. These conditions result in potential power densities of EFD processes that are many orders of magnitude higher than those of Van de Graaff generators.

The electrically charged particles which enter the conversion section can be generated in many ways; for instance, by a corona discharge method, as illustrated in Fig. 1, whereby ions of one polarity are introduced into the working medium. These ions migrate toward the attractor electrode; however, they cannot reach the attractor electrode due to the high speed of the working medium. Since the ions do not come in contact with the attractor electrode, the attractor current is zero. Hence, the sustaining of the electrical field between attractor and corona electrode, which produces the ions of one polarity, does not require an electric power input; instead, the total pressure of the working gas is decreased while passing through the attractor-corona electrode configuration. In this sense, fluid dynamic power, rather than electrical power, is consumed for the generation of the ions. The power required for

producing one ion is a minute fraction, on the order of 10^{-4}, of the useful electrical power output generated per ion, since the ionization voltage is around 20 volts and the generator output potential is several hundred thousand volts. Theoretical and experimental investigations conducted in the Aerospace Research Laboratories (ARL) have shown that corona discharge into high speed gas flows can produce ion current densities which are compatible with the electric current requirements and the scaling characteristics of EFD generators.

III. DERIVATION OF FUNDAMENTAL ELECTROFLUIDDYNAMIC RELATIONSHIPS

In an EFD energy conversion channel, as illustrated schematically in Fig. 2, a gaseous working medium flows through the conversion section with a velocity $\vec{v}(x,r)$. The working medium contains ions or charged colloids of one polarity, which, at the entrance to the conversion section, are confined to a fluid core of the radius R_i. The unipolar ions or charged colloids constitute a space charge field in the conversion section having an electric charge density $\rho_e(r,x)$.

Between the grounded entrance electrode and the collector electrode, a potential difference V exists, which, in the absence of space charges, produces in the conversion section an electrostatic field of constant field strength

$$E_x(x,r) \equiv \frac{V}{h} \; ; \quad E_r(x,r) \equiv 0$$

The space charge cloud which is produced within the conversion section by the ions or charged colloids of one polarity modifies the electrostatic field in accordance with the following relationships:

$$\operatorname{div} \vec{E}(x,r) = \frac{\rho_e(x,r)}{\varepsilon} \qquad \text{- - - - - - - - - - - [1]}$$

and

$$\int_0^h E_x(x,r)\,dx = V \qquad \text{- - - - - - - - - - - [2]}$$

Whereby the integral has the same value for any radius r within the conversion section.

The resulting electrostatic field in the conversion section has radial field components $E_r(x,r)$. The axial field components $E_x(x,r)$ near the ground electrode are larger, near the collector electrode are smaller, than V/h.

Throughout the electrostatic field, the conditions must be satisfied

that the field strength is sufficiently below the breakdown field strength E_b

$$E(x,r) < E_b \qquad \text{- - - - - - - - - - - [3]}$$

For example, in dry air of one atm, the value of E_b is approximately 30,000 volt/cm. The breakdown field strength increases about proportionally with increasing pressure.

$$E = \frac{P}{P_N} E_{b,N} \qquad \text{- - - - - - - - - - - [4]}$$

where subscript N indicates normal atmospheric conditions. The range of P/P_N for which Eq [4] is valid depends on the nature of the gas. For air, deviations from Eq [4] occur for $P/P_N > 30$. The breakdown field strength is also influenced by the field geometry, the potential difference, electrode material and others. Therefore, Eq [4] should only be considered as an approximation for determining trends.

The viscous coupling between the working medium and the charged particles gives rise to a drift velocity $\vec{v}_d(x,r)$ of the charged particles relative to the working medium.

$$\vec{v}_d(x,r) = k \cdot \vec{E}(x,r) \qquad \text{- - - - - - - - - - - [5]}$$

where k is the mobility of the ions or charged colloids. The resulting transport velocity $\vec{v}_e(x,r)$ is as follows:

$$\vec{v}_e(x,r) = \vec{v}(x,r) + \vec{v}_d(x,r)$$

or

$$\vec{v}_e(x,r) = \vec{v}(x,r) + k\vec{E}(x,r) \qquad \text{- - - - - - - - - - - [6]}$$

As the pressure increases, the ion mobility decreases approximately

as follows:

$$k = k_N \sqrt{\frac{P_N}{P}} \qquad \text{- - - - - - - - - - [7]}$$

Within the range of values of P_N/P for which, with good approximation, Eq [4] is valid, Eq [7] also represents a good approximation. For charged colloids of very small diameter, below about 2/100 micron, Eq [7] can also be employed. For larger colloids, however, strong deviations occur, and for particles above 2/10 micron in diameter, Stokes' law becomes valid, which means that the mobility is essentially independent of the pressure (Ref. 2).

In connection with the field of transport velocity $\vec{v}_e(x,r)$, the following two conditions must be satisfied:

a) Throughout the conversion section, the axial component of the transport velocity should be larger than zero.

$$v_{e,x}(x,r) > 0$$

or

$$v_x(x,r) > kE_x(x,r) \qquad \text{- - - - - - - - - - [8]}$$

b) The radial component of the electrostatic field $E_r(x,r)$ generates (according to Eq [6]) a radial component of the transport velocity $v_{e,r}(x,r)$ which means that the ions or charged colloids drift radially outward from the radius R_i to the radius R_o. The borderline of the cloud of ions or charged colloids within the conversion section may be defined as $r_e(x)$, which is indicated as a dotted line in Fig. 2. It is necessary that the charged particles reach the radius R_o inside the collector electrode rather than inside the dielectric duct of the energy conversion section, which can be expressed as follows:

$$r_e(h) \leq R_o \qquad \text{- - - - - - - - - - [9]}$$

If this condition is not satisfied, charge depositions occur at the dielectric walls of the conversion section. This would result in surface field breakdown under the high field strength conditions which are obtained with pressurized working media.

The fields of charge density $\rho_e(x,r)$ and transport velocity $v_e(x,r)$ determine the field of electric current density $i(x,r)$ in the conversion section, as follows:

$$\vec{i}(x,r) = \rho_e(x,r) \cdot \vec{v}_e(x,r)$$

or

$$\vec{i}(x,r) = \left(\vec{v}(x,r) + k\vec{E}(x,r)\right)\varepsilon \cdot \text{div } E(x,r) \qquad \text{- - - - - - - - - - - [10]}$$

The total electric current I can be expressed as follows:

$$I = 2\pi\varepsilon_o \int^{R_o} \left(v_x(x,r) - kE_x(x,r)\right) \text{div } \vec{E}(x,r) r dr \qquad \text{- - - - - - [11]}$$

whereby the value of I is the same for any value of X within the conversion section.

In selecting a design point, one will generally assume that the highest electric field strength occurring within the conversion section should not have an unnecessarily large margin from breakdown field strength. The electrical potential V (Eq [2]) and electric current I (Eq [11]), at design point, determine the outer load R as follows:

$$R = \frac{V_{des.}}{I_{des.}} \qquad \text{- - - - - - - - - - - [12]}$$

The above considerations can be employed to obtain numerical solutions for the axisymmetrical three-dimensional case when the geometry of the conversion section and the boundary conditions are specified. The solutions,

however, cannot be given in explicit form, except when $R_o >> h$ whereby radial components and end effects become insignificant, and the problem is essentially of one-dimensional nature (Ref. 1).

The above given equations will now be utilized for the derivation of scaling characteristics.

a) Geometric scaling: A specific geometric configuration of a conversion section (for example, one similar to that shown in Fig. 2) may be given. The electrical performance obtainable with this configuration may be an electric current I and an electric potential V which result from the fields $\vec{v}(x,r)$, $\vec{v}_d(x,r)$, $\vec{v}_e(x,r)$, $\vec{E}(x,r)$, and $\vec{i}(x,r)$ under the restricting conditions expressed by Eqs [3], [8] and [9]. It will now be determined in which manner the electrical performance and the fields in the conversion section change when all linear dimensions of the conversion section are scaled up by a factor α, while the velocity and the physical properties of the working medium, such as dielectric strength, mobility, pressure, mass density, etc., remain unchanged. The unchanged velocity condition of the working medium in the scaled-up conversion section can be expressed as follows: (The values in the scaled-up conversion section are indicated by primes.)

$$\vec{v}'(\alpha \cdot x, \alpha \cdot r) = \vec{v}(x,r) \qquad \text{- - - - - - - - - - [13]}$$

In order to satisfy, in the scaled-up conversion section, the electric field breakdown limitation in the same manner as in the original conversion section, the following equation must be satisfied:

$$\vec{E}'(\alpha \cdot x, \alpha \cdot r) = \vec{E}(x,r) \qquad \text{- - - - - - - - - - [14]}$$

Since, according to Eq [6]

$$\vec{v}_e'(\alpha \cdot x, \alpha \cdot r) = \vec{v}'(\alpha \cdot x, \alpha \cdot r) + k\vec{E}'(\alpha \cdot x, \alpha \cdot r)$$

it follows from Eqs [13] and [14] that

$$\vec{v}_e'(\alpha \cdot x, \alpha \cdot r) = \vec{v}_e(x,r) \qquad \text{- - - - - - - - - - - [15]}$$

Also, from Eq [14] it follows that

$$\text{div } E'(\alpha \cdot x, \alpha \cdot r) = \frac{1}{\alpha} \text{ div } \vec{E}(x,r) \qquad \text{- - - - - - - - - - - [16]}$$

The current density i which is, according to Eq [10], as follows:

$$\vec{i}'(\alpha \cdot x, \alpha \cdot r) = v_e'(\alpha \cdot x, \alpha \cdot r) \, \varepsilon \text{ div } \vec{E}'(\alpha \cdot x, \alpha \cdot r)$$

can be written, according to Eqs [15] and [16]

$$\vec{i}'(\alpha \cdot x, \alpha \cdot r) = \vec{v}_e(x,r) \, \frac{\varepsilon}{\alpha} \text{ div } \vec{E}(x,r)$$

or

$$\vec{i}'(\alpha \cdot x, \alpha \cdot r) = \frac{1}{\alpha} \vec{i}(x,r) \qquad \text{- - - - - - - - - - - [17]}$$

From Eqs [11] and [17], it follows that

$$I' = \alpha \cdot I \qquad \text{- - - - - - - - - - - [18]}$$

and from Eqs [2] and [14] that

$$V' = \alpha V \qquad \text{- - - - - - - - - - - [19]}$$

Summarizing, the geometric scaling characteristics under conditions of unchanged velocity and unchanged physical properties of the working medium are as follows:

Electrical charge density ρ_e and current density i are inversely proportional to the scaling factor α, the electric power output per area does not change with scaling, the total electrical current I and electrical

potential V increase proportionally with the scaling factor α .

b) - Effects of changing the pressure level: A specific geometric configuration of a conversion section may be given. It will be determined in which manner the electrical performance and fields in the conversion section change when the pressure level is changed from atmospheric pressure P_N to a higher pressure P , while the velocity or Mach number of the working medium are unchanged. (The values under high pressure conditions are indicated by primes.)

From Eq [4], it follows that

$$\vec{E}'(x,r) = \frac{P}{P_N} \vec{E}_{b,N}(x,r) \qquad \text{- - - - - - - - - - [20]}$$

and hence $$V' = \frac{P}{P_N} V \qquad \text{- - - - - - - - - - [21]}$$

also $$\rho_e'(x,r) = \varepsilon \operatorname{div} \vec{E}'(x,r) = \frac{P}{P_N} \varepsilon \operatorname{div} \vec{E}(x,r) \qquad \text{- - - - - - - - [22]}$$

From Eqs [6] and [7], it follows that

$$\vec{v}_e'(x,r) = \vec{v}(x,r) + \frac{P_N}{P} k_N \vec{E}'(x,r)$$

employing Eq [20] results in

$$\vec{v}_e'(x,r) = \vec{v}(x,r) + k_N \cdot \vec{E}_N(x,r)$$

or

$$\vec{v}_e'(x,r) = \vec{v}_e(x,r) \qquad \text{- - - - - - - - - - [23]}$$

The electric current density $\vec{i}'$ is, according to Eq [1], as follows:

$$\vec{i}'(x,r) = \vec{v}_e'(x,r) \cdot \varepsilon \operatorname{div} \vec{E}'(x,r); \qquad \text{(with Eq [22] and [23])}$$

$$\vec{i}'(x,r) = \vec{v}_e(x,r) \cdot \varepsilon \frac{P}{P_N} \operatorname{div} \vec{E}_N(x,r)$$

or $$\vec{i}'(x,r) = \vec{i}(x,r) \frac{P}{P_N} \text{ ; also } I' = I \cdot \frac{P}{P_N} \qquad \text{- - - - - - - - - - [24]}$$

Summarizing, the effects of increasing pressure level from atmospheric pressure P_N to a pressure P with unchanged geometry of the conversion section, unchanged velocity of the working medium, and with the conditions given for Eqs [4] and [7], are as follows: The fields of charge density ρ_e, electric current density $\vec{i}$, and electric field strength $\vec{E}$, also the electric current I and the electric potential V increase proportionally with P/P_N, while the fields of transport velocity $\vec{v}_e$ and drift velocity $\vec{v}_d$ of ions or charged colloids of very small diameter (below approximately 2/100 micron) are not changed by pressure increase. The electric power output $I \cdot V$ increases with the square of P/P_N.

In the following, the conversion efficiency from fluid dynamic into electrical energy and other EFD performance parameters will be derived. To study trends, it is desirable to have explicit relationships. Hence, the derivations will be based on the case where $R_o \gg h$, which corresponds to plane parallel electrode configurations, allowing a one-dimensional treatment. Axisymmetrical three-dimensional configurations, such as those shown in Figs. 1 and 2, have also been investigated in ARL. However, these configurations cannot be treated in a rigorous and explicit manner (except for the limiting case $R_o \ll h$ (Ref. 5)); rather, numerical approximations must be employed. Results of ARL's numerical investigations showed that optimum configurations have length diameter ratios ranging between approximately 2 and 3. The corresponding power density is about two times higher than that of a plane parallel electrode configuration.

For the one-dimensional plane parallel case, the current density i (see Eq [10]) can be written as follows:

$$i = - \varepsilon \frac{dE}{dx} (v - k \cdot E_x) \qquad \text{- - - - - - - - - - [25]}$$

Integration of this equation from the entrance to the conversion section $x = o$ to a point x between entrance and collector electrode results in the following expression for E_x :

$$i \cdot x = -\varepsilon v(E_1 - E_x) = \frac{\varepsilon k}{2}(E_1^2 - E_x^2) \qquad \text{- - - - - - - - - - [26]}$$

Let the field strength at the exit of the conversion duct (at $x = h$) be γE_1 , where γ is to be determined for the condition of maximum power. Substituting this boundary condition into Eq [3] results in:

$$\frac{i \cdot h}{v} = \varepsilon E_1(\gamma - 1) - \frac{\varepsilon k E_1^2}{2v}(\gamma^2 - 1) \qquad \text{- - - - - - - - - - [27]}$$

The factor $\frac{kE_1}{2v}$ represents one half of the ratio of the ion drift speed at the entrance electrode $v_{d,1}$ to the speed of the working substance v . This ratio will be denoted by (λ) .

$$\lambda = \frac{v_{d,1}}{2v} = \frac{kE_1}{2v} \qquad \text{- - - - - - - - - - [28]}$$

Substituting the value of i given by Eq [27] into Eq [26] results in

$$\frac{\varepsilon v E_1}{h}(1 - \gamma)[1 - \lambda(1+\gamma)]x = -\varepsilon v(E_x - E_1) + \frac{\varepsilon k}{2}(E_x^2 - E_1^2)$$

Solving for $\frac{x}{h}$ gives:

$$\frac{x}{h} = \frac{(1 - \frac{E_x}{E_1})[1 - \lambda(1 + \frac{E_x}{E_1})]}{(1 - \gamma)[1 - \lambda(1 + \gamma)]} \qquad \text{- - - - - - - - - - [29]}$$

The electric power output of the generator per unit electrode area is

$$L_e = iV = i \int_{x=o}^{x=h} E_x dx = ihE_1 \int_{x/h=o}^{x/h=1} \frac{E_x}{E_1} d(\frac{x}{h}) \qquad \text{- - - [30]}$$

Solving for $d(\frac{x}{h})$ in terms of $\frac{E_x}{E_1}$ and $d(\frac{E_x}{E_1})$ from Eq [29] and substituting

into Eq [30] gives:

$$L_e = ihE_1 \int_1^{\gamma} \frac{E_x}{E_1} \cdot \frac{(2\lambda\frac{E_x}{E_1} - 1)}{(1 - \gamma)[1 - \lambda(1 + \gamma)]} d(\frac{E_x}{E_1})$$

Substituting the value of i from Eq [27] and performing the integration results in:

$$L_e = \frac{\varepsilon v E_1^2}{2} \left[\frac{4\lambda}{3}(\gamma^3 - 1) - (\gamma^2 - 1)\right] \quad \text{- - - - - - - - - - - [31]}$$

This expression is a maximum for $\gamma = o$, that is $E_h = o$. Eq [31] becomes:

$$L_{e(max)} = \frac{\varepsilon v E_1^2}{2} \left| 1 - \frac{4\lambda}{3} \right|$$

$$= \frac{\varepsilon k E_1^3}{2} \left| \frac{3 - 4\lambda}{6\lambda} \right|$$

Substituting into the above equation the breakdown field strength E_b results in:

$$L_{e(max)} = \frac{\varepsilon k E_b^3}{2} \left| \frac{3 - 4\lambda}{6} \right| = \frac{\varepsilon v E_b^2}{2} \left| 1 - \frac{4\lambda}{3} \right| \quad \text{- - - - - - [32]}$$

and, from Eq [7],

$$L_e = \varepsilon k_N E_{b,N}^3 (\frac{P}{P_N})^2 (\frac{1 - 4\lambda/3}{4\lambda}) \quad \text{- - - - - - - - - - [32a]}$$

The speed of the working substance v between the electrodes follows from Eq [28] and by the introduction of E_b,

$$v = kE_b/2\lambda$$

or, employing Eqs [4] and [7]

$$v = \frac{k_N E_{b,N}}{2\lambda}$$

Introduction of γ being zero into Eq [29] results in the following

expression of field strength E_x :

$$\frac{x}{h} = \left(1 - \frac{E_x}{E_1}\right)\left(1 - \frac{\lambda}{1 - \lambda}\frac{E_x}{E_1}\right) \qquad \text{- - --- - - - - - - [33]}$$

In Fig. 3, $\frac{E_x}{E_1}$ is plotted as a function of $\frac{x}{h}$ curve parameter being λ . It can be seen that $\frac{E_x}{E_1}$ becomes a linear function of $\frac{x}{h}$ for $\lambda = 0$, which corresponds to the condition of constant electric charge density between the electrodes. For λ approaching 1/2, the speed of the working substance approaches the drift speed of the ions at the entrance electrode. This results in a large non-uniformity of the transport speed of the ions (approaching zero at the entrance electrode) and consequently in a strongly non-uniform charge density distribution between the electrodes with a steeply increasing charge density at the entrance electrode.

The total pressure drop ΔP_{Tot} in an incompressible working substance between the inlet and exit electrodes due to the interactions of the electrostatic field with the electric charges in the working substance is:

$$\Delta P_{Tot} = \int_{x=o}^{x=h} \rho_e(x)E(x)dx = \epsilon \int_{x=o}^{x=h} \frac{dE(x)}{dx} E(x)dx$$

Introducing the electric power maximizing condition $(E(h) = 0)$ and the field breakdown limitation, that $E_1 = E_b$, results in the following expression:

$$\Delta P_{Tot} = \frac{\epsilon}{2} E_b^2 \qquad \text{- - - - - - - - - - [34]}$$

For normal atmospheric conditions, E_b may be approximately 3×10^6 volt/m and hence ΔP_{Tot} would be 40 Newtons/m^2, or 4×10^{-4} atmospheres. The total pressure ratio $\frac{P_{t,1}}{P_{t,h}}$ would be 1.0004. Since E_b increases proportionally

with increasing pressure up to about 30 atm, ΔP_{Tot} at 30 atm would be about 3.6×10^{-1} atm and the pressure ratio would be 1.012.

The fluid dynamic power L_f per unit flow cross section extracted from the working substance by the interaction of the ions with the electric field is as follows:

$$L_f = \Delta P_{Tot} \cdot v = \epsilon \frac{E_b^2}{2} v = \frac{L_e}{1 - 4\lambda/3} \qquad \text{- - - - - - - - - - - [35]}$$

The fluid dynamic power loss L_d per unit electrode area, which is caused by the friction between the flow and the confining walls and grid electrodes, can be written as follows:

$$L_d = \xi \rho \frac{v^3}{2} \qquad \text{- - - - - - - - - - - [36]}$$

where ξ is an overall drag loss coefficient.

The electrical stage efficiency of conversion of fluid dynamic energy into electrical energy η_{st}, whereby the exit kinetic fluid power is not considered a loss, is as follows:

$$\eta_{st} = \frac{L_e}{L_f + L_d} \qquad \text{- - - - - - - - - - - [37]}$$

Introduction of Eqs [36], [35], [32] and [28] into Eq [37] results in the following expression for the efficiency η_{st} :

$$\eta_{st} = \frac{1 - 4\lambda/3}{1 + \xi \frac{\rho k^2}{4\epsilon\lambda^2}} \qquad \text{- - - - - - - - - - - [38]}$$

The dimensionless term $\frac{4\epsilon}{\rho k^2}$ will be denoted by K ; the value of K depends on the nature and the state of the working substance and can be considered as a figure of merit in analogy to the figure of merit in magneto-fluiddynamics, which is the ratio of conductivity to mass density. From

Eq [7], $K = K_N \frac{P}{P_N}$.

Introducing the term K into the efficiency expression results in:

$$\eta_{st} = \frac{1 - 4\lambda/3}{1 + \frac{\xi}{K\lambda^2}} \qquad \text{- - - - - - - - - - - [38a]}$$

For air under normal atmospheric conditions K is determined as follows:

normal mass density, $\rho_N = 1.22 \frac{kg}{m^3}$

$$\varepsilon = 8.9 \times 10^{-12} \frac{\text{Amp-sec}}{\text{volt-m}}$$

$$k_N = 1.6 \times 10^{-4} \frac{m}{\text{volt-sec}}$$

and hence $K_N = 1.13 \times 10^{-3}$

Substantially higher values of K are obtainable by employing other gases; some relative values, taking air as unity, are

H_2	1.1	C_2H_2	4.6	HCL	7.2
CO_2	2.4	SO_2	6.9	NH_3	13.9

In Fig. 4, the conversion efficiency η_{st} is plotted versus λ , curve parameter is ξ/K . It can be seen from Fig. 4 that favorable efficiencies can be expected only for very small values of ξ/K . The top curve $(\xi/K = 0)$ corresponds to the limiting case of no aerodynamic losses due to channel walls or electrodes; the losses are due solely to relative motion or slip between the charges and the neutrals.

To obtain a perspective of obtainable efficiencies, let us compare three different cases, using:

a) Air at normal atmospheric values of pressure and temperature;

b) Like a) but 20 atm pressure;

c) Ammonia at 5 atm pressure.

In all three cases, a drag loss coefficient of 3% will be taken. In the first case, the stage efficiency η_{st} is on the order of 0.3%. In the second case, for air under 20 atm pressure, the efficiency is about 5%. While for the third case, using ammonia gas under 5 atm, the efficiency is about 16%. Much higher figures of merit than that of ammonia would be required to obtain 50% stage efficiency. Investigations are required to determine whether special gas or vapor mixtures may be realized to satisfy these conditions.

To achieve stage efficiency values more nearly equal to 100%, the application of colloids as charge carriers is necessary, as well as the development of mixtures of gases having the combined physical properties of low mass density and high dielectric strength. Fig. 5 illustrates for air the influence of colloid diameter upon mobility and drift velocity under the influence of electric fields corresponding to breakdown values (Ref. 2). Fig. 6 is a similar figure for pure hydrogen. It is seen for even quite small colloid diameters of .01 micron that the drift velocity is reduced to almost 1% of that for ions. Hence, the relative motion or slip for E, a field strength nearly equal to E_b, may be made negligible.

For the case of λ approaching zero, it is useful to express K as $\frac{\varepsilon E_1^2}{\rho \lambda^2 v^2}$ and the term $\frac{\xi}{K\lambda^2}$ in Eq [38a] becomes

$$\frac{\xi \frac{\rho v^2}{2}}{\frac{\varepsilon E_1^2}{2}} \quad \text{or} \quad \frac{\xi \rho_N v^2}{\varepsilon E_{1,N}^2} \cdot \frac{P_N}{P}$$

Eq [38a] then becomes for $E_{1,N}$ taken as equal to $E_{b,N}$

$$\eta_{st}\ (\lambda = 0) = \frac{1}{1 + \frac{\xi \rho_N v^2}{\varepsilon E_{b,N}^2} \cdot \frac{P_N}{P}} \qquad \text{- - - - - - - - - - [39]}$$

This equation is plotted in Fig. 7 for air as the working medium; also plotted is the electrical power density, which becomes:

$$L_e = \varepsilon \cdot \frac{E_{b,N}^2}{2} \cdot \left(\frac{P}{P_N}\right)^2 \cdot v \qquad \text{- - - - - - - - - - - [40]}$$

Performance values for hydrogen with suitable additives in order to obtain the same breakdown value as that of air are shown in Fig. 8.

The most important EFD performance characteristics are summarized in Fig. 9, and, in Figure 10, total pressure ratio, stage efficiency and useful specific power output for three different cases are given. In the first case, a working gas under atmospheric pressure, containing unipolar ions and having a mass density and dielectric strength similar to that of air, is assumed. A favorable drag loss coefficient of 3% is taken for the conversion section and electrodes. This value will be used also for the other two cases. The resulting performance values of the first case are by many orders of magnitude too poor for any normally conceivable application. Discouraging performance characteristics such as these may be the reason for the generally unexplored state of EFD processes.

In the second case, by pressurization of the same working gas to about 20 atm, and by employment of charged colloids rather than ions, about a thousand-fold performance improvement is obtained; i.e., specific electric power outputs of about 400 watts/cm^2 and stage efficiencies of about 40% become potentially feasible. The third case is based upon a light molecular weight working gas, such as hydrogen, under the same pressure as in the second case. It is assumed that by suitable additives the same dielectric strength as that of air can be obtained. In this case, the performance is improved further to about 750 watts/cm^2 and to a stage efficiency of about 74%.

These results show that very favorable power densities and stage efficiencies become possible with viscous EFD energy conversion processes which (a) employ electrode and channel configurations with aerodynamic shapes having a very small drag loss, and (b) utilize pressurized working gases of low molecular weight and high dielectric strength containing charged colloids of one polarity. However, as can be seen from the bottom row of Fig. 10, the pressure ratios are very low, which makes it difficult to employ a one-stage viscous EFD process directly in a closed thermodynamic cycle. These problems will be treated in the next section.

IV. THERMODYNAMIC ASPECTS OF EFD ENERGY CONVERSION PROCESSES

The most important potential applications of EFD processes are those in the field of energy conversion from heat into electrical energy. In such processes, heat is transformed into fluid dynamic energy which in turn is directly transformed by the EFD process into electrical energy. The waste heat, that is, the heat energy not transformed into electrical energy, must be rejected, for example, by a space radiator.

The efficiency η of a complete cycle from heat into electrical energy can be expressed as follows:

$$\eta = \eta_{therm}\, \eta_{el}$$

where η_{therm} is the efficiency of the conversion from heat into fluid dynamic energy of the EFD working medium, and η_{el} is the efficiency of the direct conversion from fluid dynamic energy into electrical energy. The distinction between η_{el} and the previously used stage efficiency η_{st} is as follows: In the definition of stage efficiency, as stated before, the kinetic energy at the exit of the conversion section is not accounted for as a loss. In a complete cycle, however, the kinetic energy L_K of the working gas leaving the conversion section is converted into static energy by a diffusor shown in Fig. 1. The diffusor loss is $L_K\,(1 - \eta_{diff})$, where η_{diff} represents the diffusor efficiency. This diffusor loss is included in the definition of η_{el} for a working substance containing charged particles as given in Fig. 11.

$$\eta_{el} = \frac{L_e}{L_e + \xi L_K + (1 - \eta_{diff}) \cdot L_K}$$

where L_e represents the electric energy output and ξ the previously mentioned loss coefficient of the conversion section. In this Fig. η_{el} is

plotted versus L_e/L_K . The curve for $\eta_{diff} = 1$ is identical with the stage efficiency. It can be seen from Fig. 11 that η_{el} generally lies considerably below the stage efficiency, particularly in the regime of low values of L_e/L_K . This condition could be improved by multistaging, whereby η_{el} would be as follows:

$$\eta_{el} = \frac{L_e/L_K}{L_e/L_K + \xi + \frac{1 - \eta_{diff}}{\text{Number of Stages}}}$$

However, multistaging will be very difficult to accomplish with working gases containing charged colloids. When, in view of these difficulties, single stage EFD processes are used, the values of L_e/L_K should not be smaller than about 0.1, as can be seen from Fig. 11. Unfortunately, this requirement limits the velocity of the working gas. From Fig. 12, it follows that for $\frac{Le}{L_K} = 0.1$, the speed of air at 20 atm would be about 100 m/sec, that of pure hydrogen at 20 atm about 230 m/sec. From a standpoint of high power output density, it would be desirable to employ higher velocities than these. Somewhat more favorable conditions can be obtained by departing from the plane parallel electrode geometry and employing, for instance, a cylindrical geometry as shown in Fig. 1 (Ref. 5). Also, when hydrogen is employed as an EFD working medium, much higher speeds than the above value of 230 m/sec will become possible, if the breakdown field strength of hydrogen is increased by adding suitable impurities. These means allow sufficient speed increases of the EFD working medium, which enables power output densities around 500 watt/cm^2, while the condition for obtaining favorable values of η_{el} , namely $\frac{Le}{L_K} \geq 0.1$, is satisfied.

The previous considerations showed that, for obtaining favorable performance characteristics of EFD energy conversion processes, very special

properties of the working substance are required. These properties may be incompatible with those required for a working medium to transform heat efficiently into fluid dynamic energy. Also, as shown previously, the pressure ratio of an efficient EFD energy conversion process is relatively low, and is therefore incompatible with the pressure ratio requirements for efficient conversion from heat into fluid dynamic energy. These difficulties can be overcome by employing two loops with two working media, one for the thermodynamic process to transform heat into fluid dynamic energy, the other for the EFD process to transform fluid dynamic energy into electrical energy (Fig. 13). In such a two-component, two-loop system, the conflicting requirements of the thermodynamic and the EFD conversion process are essentially uncoupled, and favorable conditions can be provided for each individual process. The EFD process may employ a light molecular weight gas of high dielectric strength, which recirculates in the pressurized EFD loop with a suitable speed through the conversion section and is kept at a proper temperature level. The thermodynamic process may use a steam, preferably of higher molecular weight than that of the EFD working medium, and may employ, for example, a super-heated Rankine cycle of high pressure ratio. Under these conditions, the energy transfer from the thermodynamic to the EFD working medium can be accomplished efficiently by mixing, because the speed difference between the working media of suitably chosen molecular weights can be kept reasonably small. For example, the speed of the thermodynamic working medium can be less than twice the speed of the EFD working medium, while the requirements of a very high expansion pressure ratio of the thermodynamic loop, and a very low pressure ratio for the EFD loop, are satisfied.

V. CONCLUSION

In summary, energy conversion by the normal viscous coupling process between ions and air at atmospheric pressure is quite unattractive, a fact which may well have discouraged interest in the past. However, as shown in this paper, striking improvements appear possible by utilizing multi-component working media, by increasing the pressure level, by employing special aero-thermodynamic cycles, and by achieving more effective coupling. The fact that these few parameter changes show such significant potential improvement in the performance is an encouraging incentive to further research. The following are typical areas where further research is indicated:

1) Free Condensation Phenomena Around Ions:

   Of particular interest is the study of the effects of physical characteristics of vapor and carrier gas, cooling rate, and other factors, upon condensation phenomena, such as nucleation around ions, droplet growth rate, droplet size, and size uniformity.

2) Geometric Parameters of the Conversion Region:

   Of major importance is the investigation of various fundamental electrode and conversion section configurations with respect to aerodynamics, electric field structure, similarity and scaling characteristics.

3) Effects of Gaseous Mixtures on Breakdown Field Strength:

   These studies should include effects of additives in light molecular weight EFD working gases, and of condensation phenomena of supersaturated vapor components around ions.

4) Trends Toward Neutralization of Space Charge Effects:

   Present investigations are based on the employment of

ions or charged particles of one polarity. If space charge neutralization could be accomplished in EFD conversion processes, current and power densities could be greatly increased.

5) Ion Generation for EFD Processes:

While present studies are mainly concerned with corona discharge, other ionization methods, such as contact, bombardment and thermal ionization, may become important possibilities in the future.

6) EFD Processes Which Allow the Generation of Alternating Current:

Although the analyses in this paper are based on the assumption of steady state conditions, many aspects of this general approach could also be applied to non-steady electrical phenomena, with the attractive possibility of generating alternating current.

7) Methods of Wall Cooling and Prevention of Electric Discharge Along the Wall of the Conversion Duct:

Since thermodynamic processes involve considerably higher temperatures than ambient, the structural insulating material of the conversion section must be kept cool for insuring high dielectric strength. Another important factor is the prevention of surface discharge along the duct walls, which points toward the investigation of various novel electro-aerodynamic approaches.

In conclusion, the field of EFD energy conversion can be considered an attractive area for exploratory research. Although technological development is premature, it is hoped that these research efforts will show the most promising pathways to applied research and development.

REFERENCES

1. Gourdine, M. C.: Power Addition and Extraction from Gas Flow by Means of Electric Wind. TR No. 32-18, Jet Propulsion Laboratory, California Institute of Technology (June 1960).

2. Decaire, Lt. J.: Phenomena of Partial Condensation Around Ions in Multi-Component Working Media for Electro-Fluid Dynamic Energy Conversion Processes. In preparation at the Aerospace Research Laboratories as a research thesis topic for the Air Force Institute of Technology.

3. Lawson, M. O., von Ohain, H. J. P. and Wattendorf, F. L.: Performance Potentialities of Direct Energy Conversion Processes Between Electrostatic and Fluid Dynamic Energy. Aerospace Research Laboratories report ARL 178 (1961).

4. Marks, A. M.: Heat Electrical Power Conversion Through the Medium of a Charged Aerosol. Patent #2,638,555 (May 1953).

5. Kahn, B. and Gourdine, M. C.: A Basic Study of Slender Channel Electrogasdynamics. Aerospace Research Laboratories report ARL 63-205 (Nov 1963).

6. Marks, A., Barreto, E., and Chu, C. K.: Charged Aerosol Energy Converter. AIAA Journal (Jan 1964).

ELECTRO-FLUID DYNAMIC POWER GENERATOR
SCHEMATIC VIEW

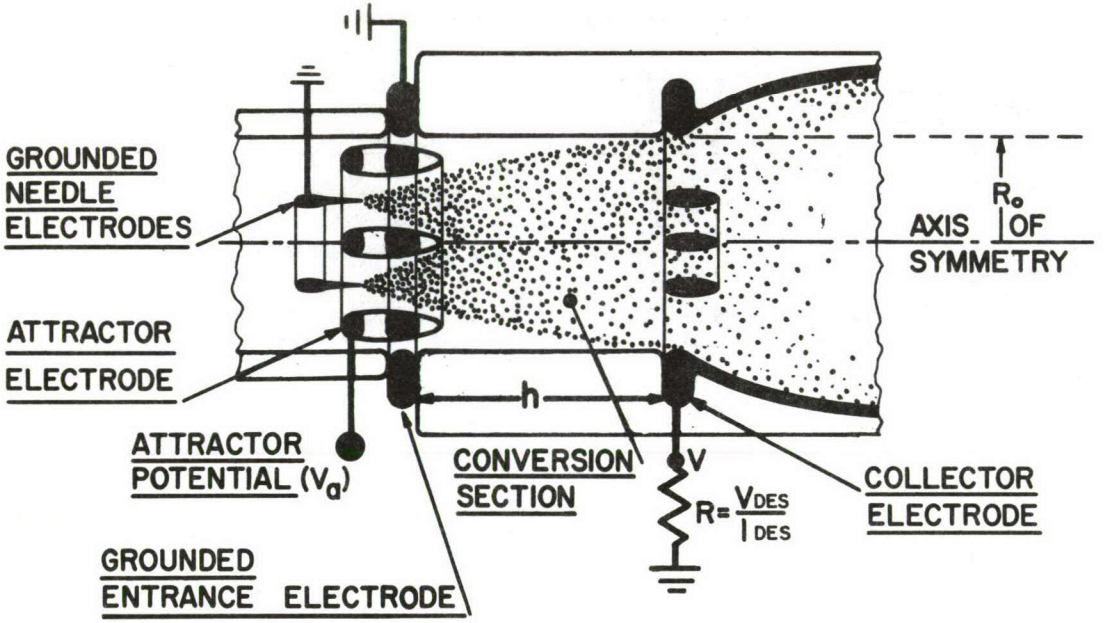


Fig. 1

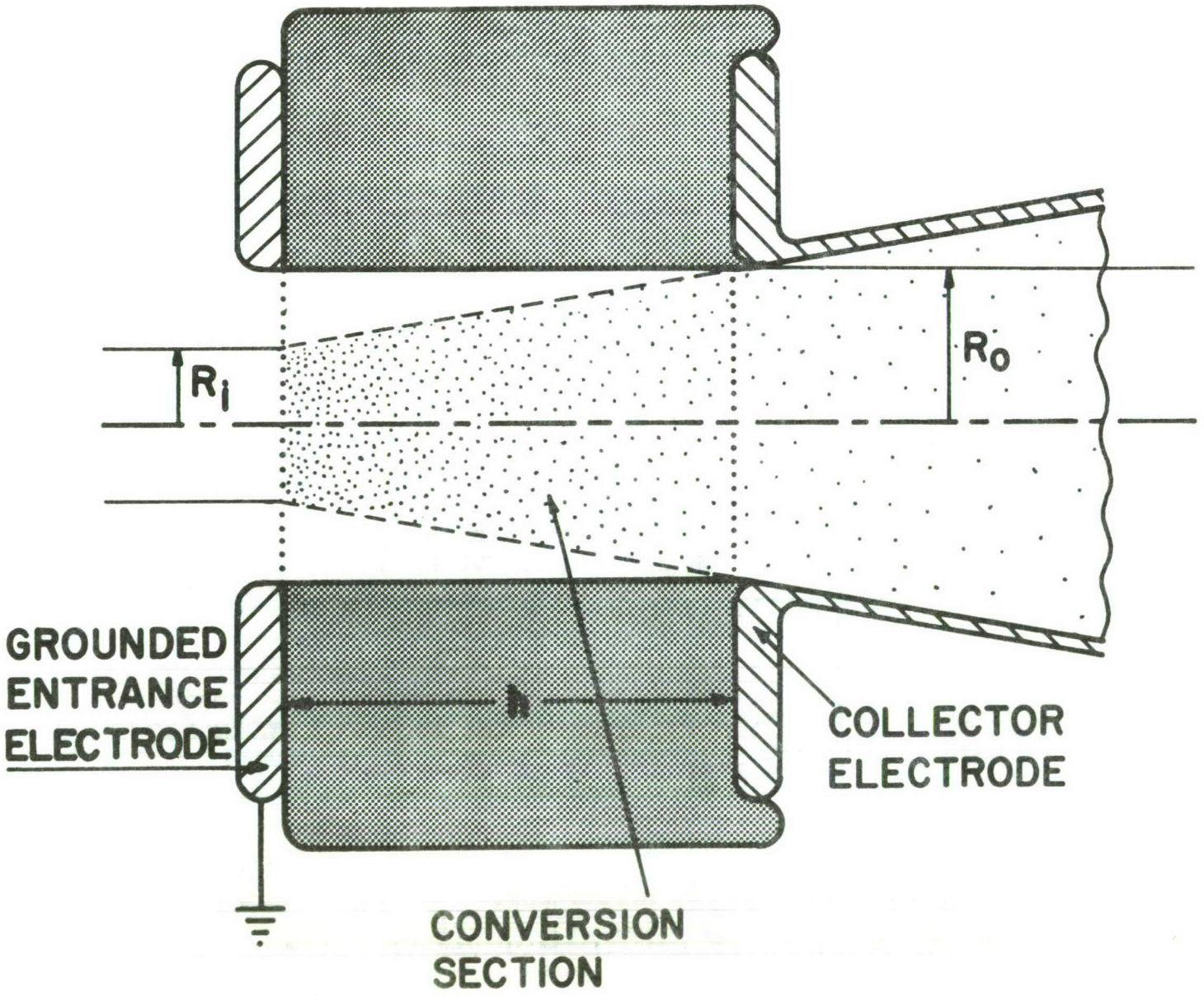

R_i
R_o
GROUNDED
ENTRANCE
ELECTRODE
COLLECTOR
ELECTRODE
CONVERSION
SECTION

Fig. 2

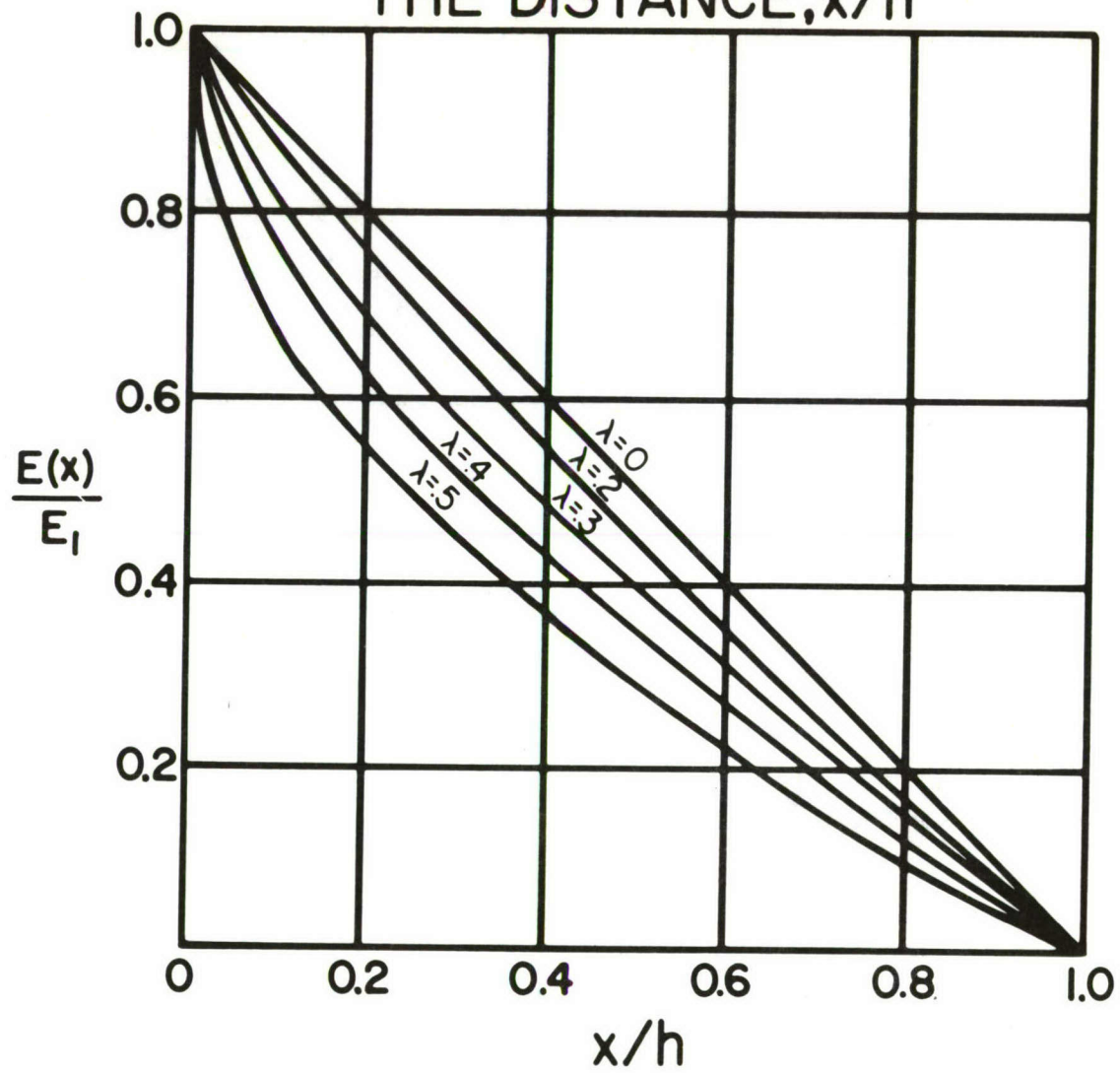
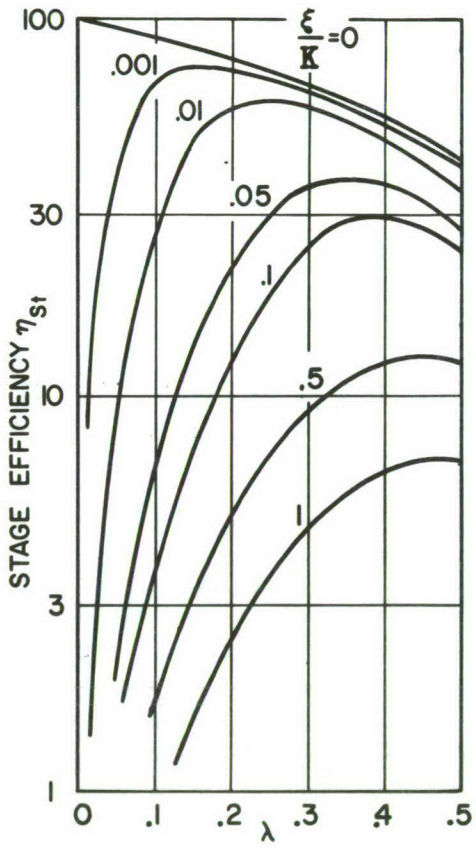

FIELD STRENGTH, E(x)/E1, AS FUNCTION OF THE DISTANCE, x/h
1.0
0.8
0.6
0.4
0.2
E(x)/E1
λ=0
λ=2
λ=3
λ=4
λ=5
0
0.2
0.4
0.6
0.8
1.0
x/h

Fig. 3



STAGE EFFICIENCY (ION)
AS A FUNCTION OF
$\lambda = \frac{kE_1}{2v}$

FOR THE PARAMETER $\frac{\xi}{K}$
WHERE $K = \frac{4\epsilon}{\rho k^2}$

$$\eta_{st} = \frac{1-\frac{4}{3}\lambda}{1+\frac{\xi}{K}\frac{1}{\lambda^2}}$$

Fig. 4

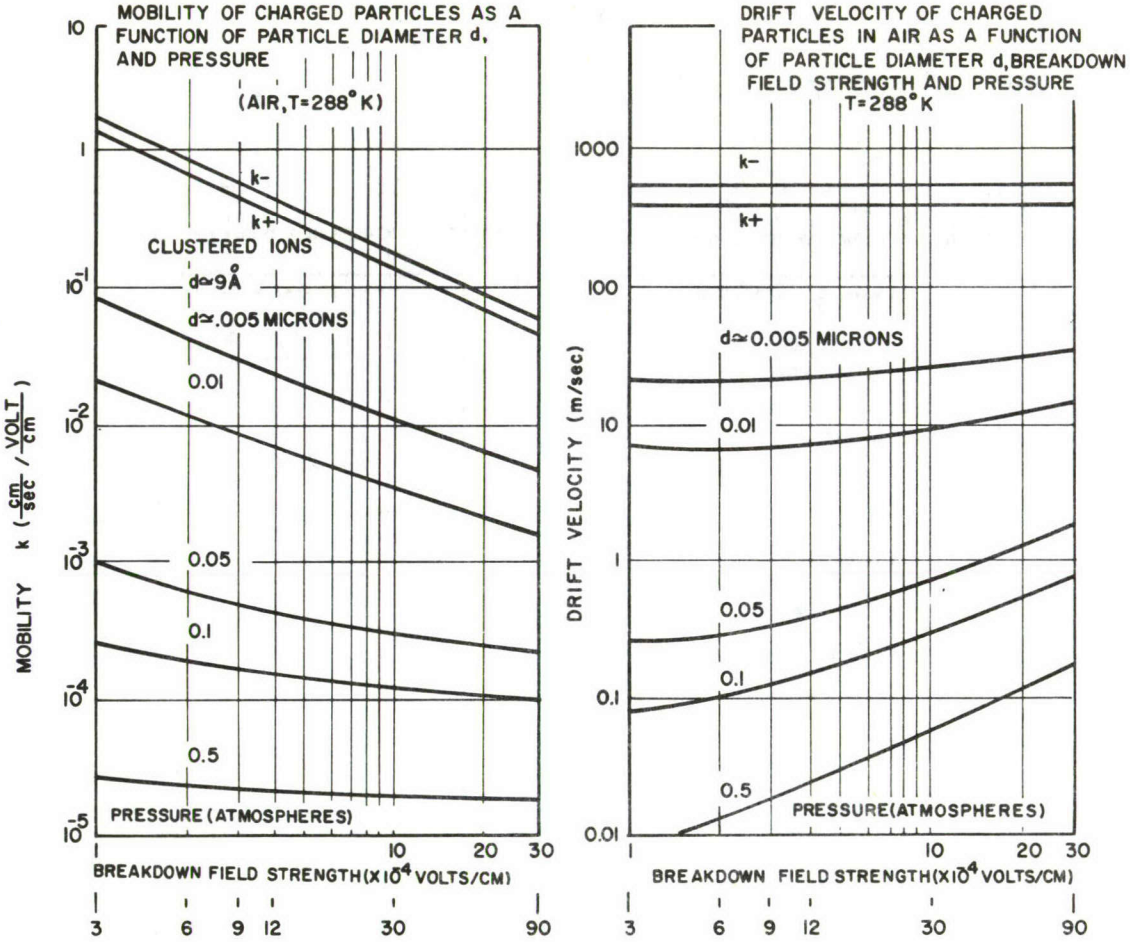

MOBILITY OF CHARGED PARTICLES AS A FUNCTION OF PARTICLE DIAMETER d, AND PRESSURE
(AIR, T=288° K)
k-
k+
CLUSTERED IONS
d≃9Å
d≃.005 MICRONS
0.01
0.05
0.1
0.5
PRESSURE (ATMOSPHERES)
MOBILITY k ($\frac{cm}{sec}$ / $\frac{VOLT}{cm}$)
BREAKDOWN FIELD STRENGTH (X10⁴ VOLTS/CM)
DRIFT VELOCITY OF CHARGED PARTICLES IN AIR AS A FUNCTION OF PARTICLE DIAMETER d, BREAKDOWN FIELD STRENGTH AND PRESSURE
T=288° K
k-
k+
d≃0.005 MICRONS
0.01
0.05
0.1
0.5
PRESSURE (ATMOSPHERES)
DRIFT VELOCITY (m/sec)
BREAKDOWN FIELD STRENGTH (X10⁴ VOLTS/CM)

Fig. 5

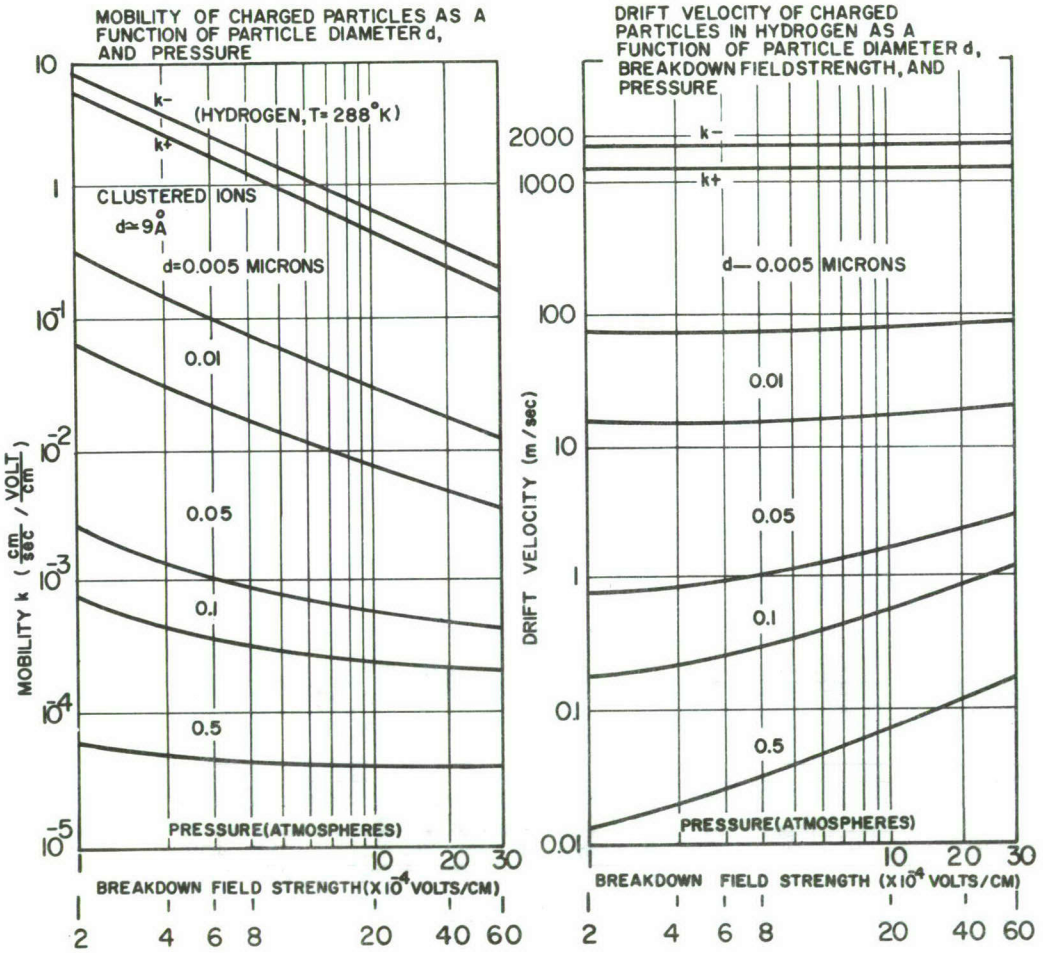
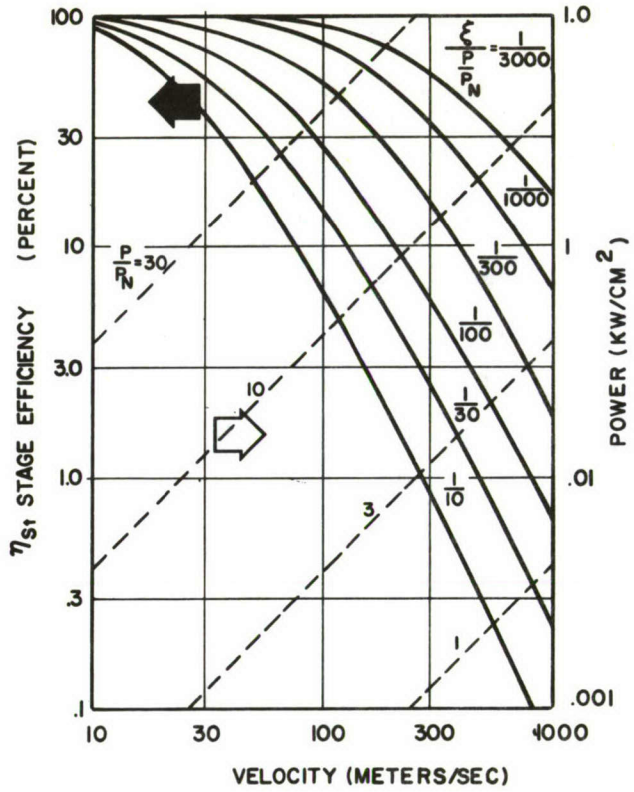

MOBILITY OF CHARGED PARTICLES AS A FUNCTION OF PARTICLE DIAMETER d, AND PRESSURE
k-
(HYDROGEN, T= 288°K)
k+
CLUSTERED IONS
d≃9Å
d=0.005 MICRONS
0.01
0.05
0.1
0.5
PRESSURE(ATMOSPHERES)
MOBILITY k (cm/sec / VOLT/cm)
BREAKDOWN FIELD STRENGTH(X10⁻⁴ VOLTS/CM)
DRIFT VELOCITY OF CHARGED PARTICLES IN HYDROGEN AS A FUNCTION OF PARTICLE DIAMETER d, BREAKDOWN FIELDSTRENGTH, AND PRESSURE
k-
k+
d— 0.005 MICRONS
0.01
0.05
0.1
0.5
PRESSURE(ATMOSPHERES)
DRIFT VELOCITY (m/sec)
BREAKDOWN FIELD STRENGTH (X10⁻⁴ VOLTS/CM)

Fig. 6



STAGE EFFICIENCY (COLLOID) AND POWER AS FUNCTIONS OF VELOCITY FOR THE CASE $\lambda = 0$ AND FOR PARAMETERS

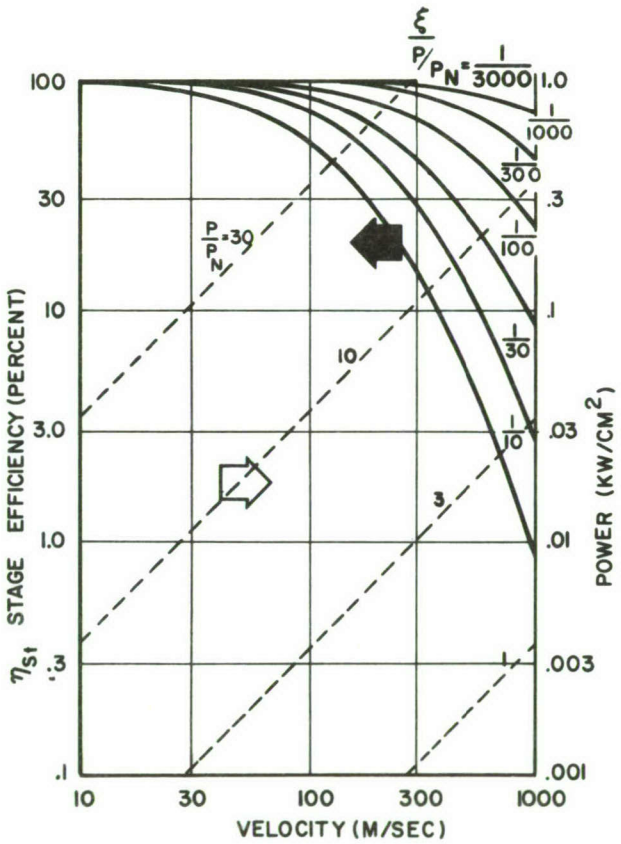
$$\frac{\xi}{\frac{P}{P_N}} \text{ AND } \frac{P}{P_N}$$

WITH AIR AS THE WORKING MEDIUM

$$\eta_{St} = \frac{1}{1 + \frac{\xi}{\frac{P}{P_N}} \frac{\rho_N v^2}{\epsilon E_{b,N}^2}}$$

$$L_e = \frac{\epsilon E_{b,N}^2}{2} \left(\frac{P}{P_N}\right) v^2$$

Fig. 7



STAGE EFFICIENCY AND POWER DENSITY AS A FUNCTION OF VELOCITY FOR THE CASE WHERE

$\lambda = 0$

WITH HYDROGEN*

$(E_{b,N} = 3 \times 10^6 \frac{V}{m})$

AS WORKING MEDIUM.

CURVE PARAMETERS

$\frac{\xi}{\frac{P}{P_N}}$ AND $\frac{P}{P_N}$

$$\eta_{St} = \frac{1}{1 + \frac{\xi \rho_N V^2}{\frac{P}{P_N} \epsilon E_{b,N}^2}}$$

$$L_e = \frac{\epsilon E_{b,N}^2}{2} \left(\frac{P}{P_N}\right)^2 V$$

*with additives

Fig. 8

PERFORMANCE CHARACTERISTICS OF VISCOUS E.F.D. ENERGY CONVERSION PROCESS

TOTAL PRESSURE DROP $\Delta P = \int_0^h \rho_e(x,r)\, E_x(x,r)\, dx$;

FOR $R_0 >> h$; AND BOUNDARY CONDITIONS: $E_{COLLECTOR} = 0$; $E_{ENTRANCE} = E_b$; $\Delta P = \frac{\epsilon E_b^2}{2}$

$\epsilon = 8.9 \cdot 10^{-12} \frac{AMP\ SEC}{VOLT\text{-}m}$; k = ION MOBILITY; h = ELECTRODE DISTANCE; v = GAS VELOCITY

	SLIP BETWEEN IONS AND WORKING MEDIUM $\lambda = \frac{DRIFT\ VELOCITY}{2 \times GAS\ VELOCITY} = \frac{k \cdot E_{ENTRANCE}}{2v}$	NEGLIGIBLE SLIP: $V_d \ll V$ (CHARGED COLLOIDS)
CURRENT DENSITY	$i = \frac{\epsilon k E_{b,N}^2}{2h} \cdot \frac{P}{P_N} \cdot \frac{1-\lambda}{\lambda}$	$i = \epsilon V \frac{E_{b,N}}{h} \frac{P}{P_N}$;
POTENTIAL	$V = \frac{E_{b,N} \cdot h}{2} \cdot \frac{P}{P_N} \frac{1-\frac{4}{3}\lambda}{1-\lambda}$;	$V = E_{b,N} \cdot \frac{h}{2} \cdot \frac{P}{P_N}$;
POWER DENSITY	$i \cdot V = L_e = \epsilon k_N E_{b,N}^3 (\frac{P}{P_N})^2 \frac{1-\frac{4}{3}\lambda}{4\lambda}$;	$i \cdot V = L_e = \epsilon E_{b,N}^2 \frac{v}{2} \cdot (\frac{P}{P_N})^2$;
STAGE EFFICIENCY	$\eta_{STAGE} = \frac{1-\frac{4}{3}\lambda}{1+\xi \rho_N k_N^2 / 4\epsilon\lambda^2 \frac{P}{P_N}}$	$\eta_{STAGE} = \frac{1}{1+\xi \frac{\rho_N v^2}{\epsilon E_{b,N}^2} \cdot \frac{P_N}{P}}$;

Fig. 9

SOME PERFORMANCE VALUES OF VISCOUS EFD PROCESSES

GIVEN CONDITIONS	TYPE OF WORKING GAS	AIR	AIR	HYDROGEN WITH ADDITIVES
	PRESSURE LEVEL	1 atm.	20 atm.	20 atm.
	MASS DENSITY	1.22 $\frac{KG}{M^3}$	24.4 $\frac{KG}{M^3}$	1.9 $\frac{KG}{M^3}$
	BREAK DOWN FIELD STRENGTH	$3 \cdot 10^4 \frac{VOLT}{cm}$	$6 \cdot 10^5 \frac{VOLT}{cm}$	$\sim 6 \cdot 10^5 \frac{VOLT}{cm}$
	TYPE OF CHARGED PARTICLES	IONS	CHARGED COLLOIDS	CHARGED COLLOIDS
	MOBILITY	$1.4 \times 10^{-4} \frac{M^2}{VOLT\ SEC.}$	$1 \times 10^{-8} \frac{M^2}{VOLT\ SEC.}$	$2.5 \times 10^{-7} \frac{M^2}{VOLT\ SEC.}$
	SLIP RATIO $\lambda = \frac{DRIFT\ VELOCITY}{2 \times GAS\ VELOCITY}$	0.5	.001	.0015
	VELOCITY OF WORKING GAS	$250 \frac{M}{SEC.}$	$250 \frac{M}{SEC.}$	$470 \frac{M}{SEC.}$
	DRAG COEFFIENCT (ξ) OF CONVERSION SECTION	$\frac{3}{100}$	$\frac{3}{100}$	$\frac{3}{100}$
RESULTING PERFORMANCE VALUES	SPECIFIC POWER OUTPUT	$.4 \frac{WATT}{cm^2}$	$400 \frac{WATT}{cm^2}$	$750 \frac{WATT}{cm^2}$
	STAGE EFFICIENCY	0.3%	41%	74%
	PRESSURE DROP ELECTRIC + DRAG	0.012atm.	.40atm.	.22atm.
	PRESSURE RATIO	1.009	1.016	1.009

Fig. 10

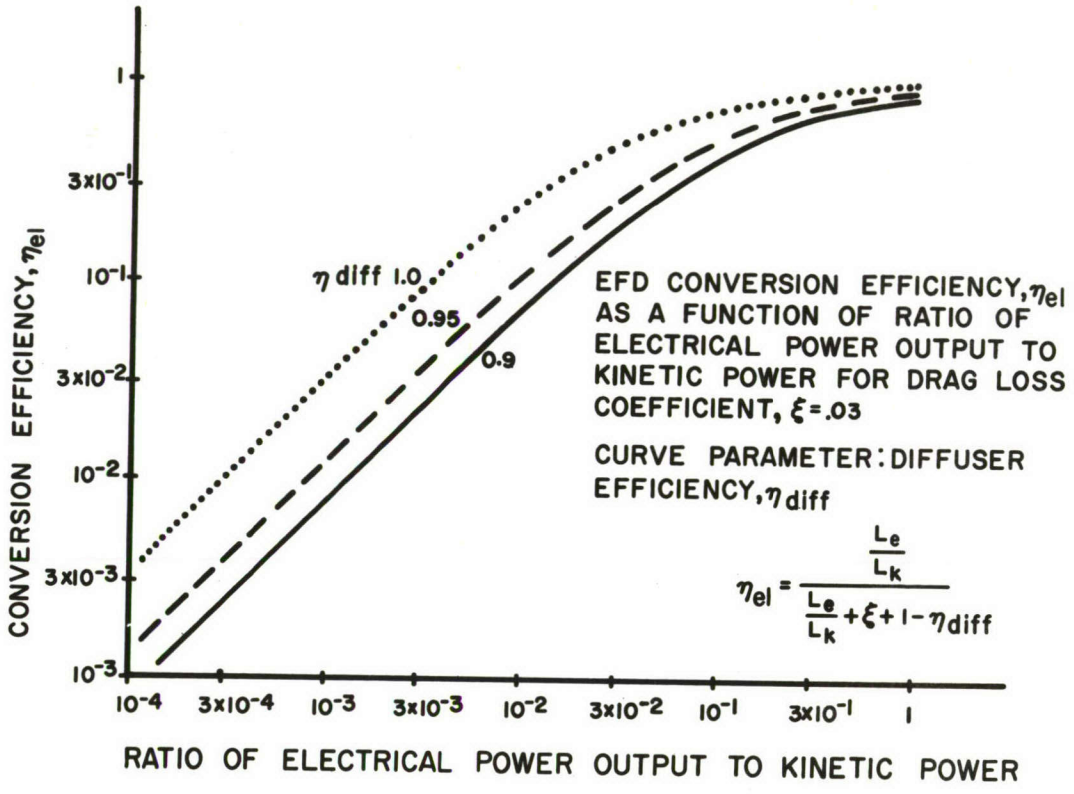

CONVERSION EFFICIENCY, η_{el}
1
3×10^{-1}
10^{-1}
3×10^{-2}
10^{-2}
3×10^{-3}
10^{-3}
η diff 1.0
0.95
0.9
EFD CONVERSION EFFICIENCY, η_{el} AS A FUNCTION OF RATIO OF ELECTRICAL POWER OUTPUT TO KINETIC POWER FOR DRAG LOSS COEFFICIENT, $\xi = .03$
CURVE PARAMETER: DIFFUSER EFFICIENCY, η_{diff}
$\eta_{el} = \frac{\frac{L_e}{L_k}}{\frac{L_e}{L_k} + \xi + 1 - \eta_{diff}}$
10^{-4} 3×10^{-4} 10^{-3} 3×10^{-3} 10^{-2} 3×10^{-2} 10^{-1} 3×10^{-1} 1
RATIO OF ELECTRICAL POWER OUTPUT TO KINETIC POWER

Fig. 11

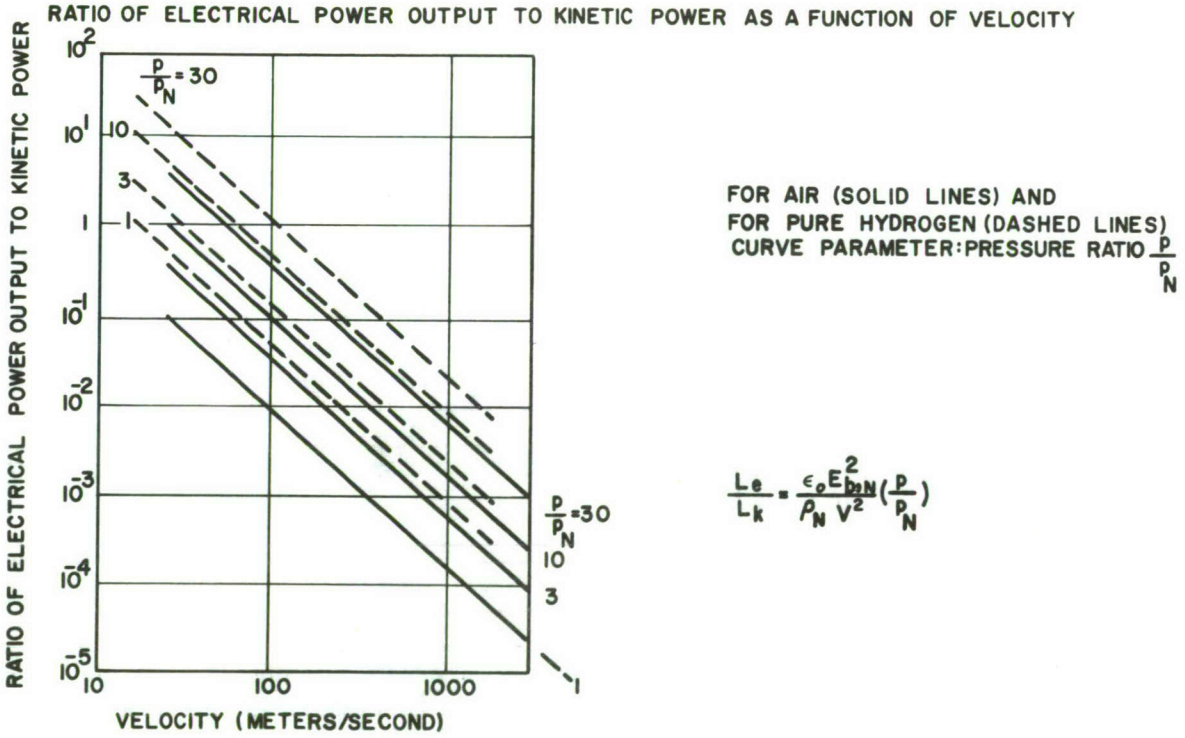

RATIO OF ELECTRICAL POWER OUTPUT TO KINETIC POWER AS A FUNCTION OF VELOCITY
RATIO OF ELECTRICAL POWER OUTPUT TO KINETIC POWER
10^2
10^1
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
$\frac{p}{p_N}$ = 30
10
3
1
$\frac{p}{p_N}$=30
10
3
1
10
100
1000
VELOCITY (METERS/SECOND)
FOR AIR (SOLID LINES) AND
FOR PURE HYDROGEN (DASHED LINES)
CURVE PARAMETER: PRESSURE RATIO $\frac{p}{p_N}$
$\frac{L_e}{L_k} = \frac{\epsilon_o E_{b_N}^2}{\rho_N V^2}\left(\frac{p}{p_N}\right)$

Fig. 12

RECIRCULATORY EFD GENERATOR,(CLOSED CYCLE SIMULATION)
SCHEMATIC VIEW

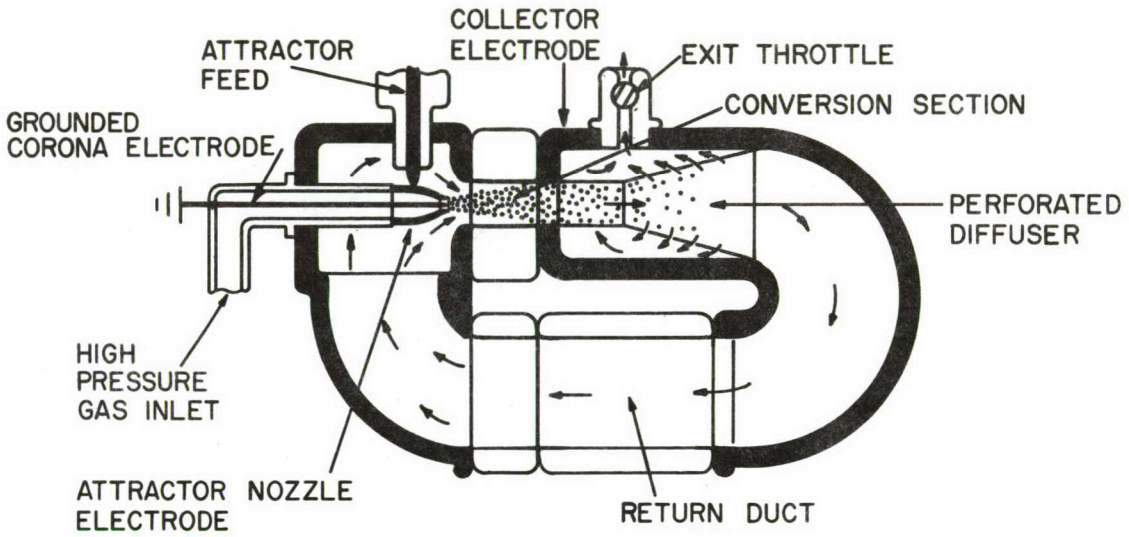


Fig. 13

ZERO GRAVITY ELECTROLYSIS CELL

BY

1st. Lt. Michael D. Rutkowski

1st. Lt. Michael A. Maxwell

26 June 1964

Project Number 6146

Air Force Flight Dynamics Laboratory
Research and Technology Division
Air Force Systems Command
Wright-Patterson Air Force Base, Ohio

ABSTRACT

(U) Zero Gravity Electrolysis Cell

The requirement for a zero gravity electrolysis cell suitable for control of space vehicle cabin atmospheres led to the initiation of a project to develop such a cell to be flight tested in an Atlas Scientific Passenger Pod. The program consisted of the fabrication and testing of a single cell, the results of which led to the design of a six cell unit. After reviewing the available literature, the reverse fuel cell type of electrolysis cell was selected because of its relative simplicity and because of extensive research done on fuel cells. A design utilizing nickel screen electrodes and an asbestos cloth - filter paper matrix was finalized utilizing a special method of feeding water to the cell in zero gravity. The electrolyte selected was potassium hydroxide because of its high ionic mobility.

The six cell unit operates at 14 volts and 12 amperes with a nominal electrode area of .230 square foot.

Problems encountered included slight hydrogen - oxygen mixing, electrolyte leakage, and foaming at the electrolyte - electrode interface. Future work will include the development of a matrix having the optimum properties of high absorbtivity and minimal gas mixing. Means of preventing electrolyte contamination of evolved gases will also be investigated.

Introduction:

A regenerating atmosphere control system for military space vehicles requires a maximum in flexibility and subsystem capability. In the event that a change in mission profile requires that the vehicle stay in orbit longer than originally planned, the atmosphere control sub-system must be able to accomodate the crew without earth-to-vehicle re-supply of oxygen. A self-sustaining air supply system would require that the oxygen be recovered from carbon dioxide and excess water vapor exhaled by the crew. Most regenerating atmosphere control systems being currently considered for aerospace vehicle application focus on the use of a carbon dioxide reduction process and a water electrolysis cell. Both processes (individually and in conjunction with one another) must reliably operate at high yield and low temperature conditions with a relatively low power requirement under conditions of zero gravity.

The AF Flight Dynamics Laboratory has recognized the need for an Air Force in-house program to build a military capability to recognize the merits and deficiencies of the state-of-the-art in atmosphere control systems. This project to design and fabricate a zero gravity electrolysis cell is specifically aimed at placing the cell on an Atlas Scientific Passenger Pod to gain significant zero gravity data in this area. The work encompassed in this paper covers the period from 1 May 63 through 1 May 64 which includes preliminary design studies through completed laboratory tests of the six cell electrolysis unit fabricated for the Passenger Pod.

Description of Process

The process of electrolysis of water produces hydrogen and oxygen by passing an electrical current through an electrolyte contained between two electrodes. The electrolyte may be of any movable ions but sulfuric acid (H_2SO_4) and potassium hydroxide (KOH) possess the highest ion mobilities and, despite their corrosive properties are the most efficient electrolytes for water electrolysis. KOH was the electrolyte used in this project.

When a potential is applied between the electrodes, the following half-cell reactions take place in a basic aqueous electrolytic medium:

$$\text{Anode:} \quad 2OH^- = \tfrac{1}{2}\, O_2 + H_2O + 2e^-$$

$$\text{Cathode:} \quad 2\, H_2O + 2e^- = H_2 + 2\, OH^-$$

$$\text{Net Reaction:} \quad H_2O \longrightarrow H_2 + \tfrac{1}{2}\, O_2$$

If this reaction were made to occur with no current flowing at constant temperature and pressure, the only energy required would be the electrical equivalent of ΔF°, the free energy of formation of H_2O. This would be the maximum work done and

$\Delta F^\circ = n\mathcal{F}E$ where

n = no. of electrons transferred

$\mathcal{F}$ = the Faraday constant

E = the voltage of the cell

Since the voltage E is the intensity factor of energy and $n\mathcal{F}$ is the capacity, the product of the two gives energy.

The free energy change (ΔF) of the reaction is equal to the change in heat content (ΔH) minus the "irreversible energy of the reaction" ($T \Delta S$) where T is the absolute temperature and ΔS is the change in entropy.

For unit activity, $\Delta F^{o} = \Delta H^{o} - T \Delta S^{o}$.

ΔH^{o} is determined from coulombmetric calculations and ΔS^{o} may be calculated from physiochemical methods. For the reaction

$$H_2O \longrightarrow H_2 + \tfrac{1}{2} O_2$$

$$\Delta H^{o}_{H_2O} = -68.317 \text{ Kcal/mol}$$

$$\Delta S^{o} = 16.716 \text{ cal/mol at } 25^{o}c$$

$$\Delta F^{o} = -56.690 \text{ Kcal/mol}$$

For this reaction, $E^{o} = \frac{\Delta F^{o}}{n \mathcal{F}} = 1.228$ volts.

This voltage is the minimum voltage at which this reaction will occur. In actual practice, higher voltage is necessary to overcome cell resistance, polarization, and electrode overvoltage. Overvoltage is a type of polarization associated with an electrode that is the difference between the voltage at which a gas is actually liberated and the cell voltage under reversible conditions. For hydrogen liberation, overvoltages range from 0.05 volt for platinum black to 1.2 volts for tin at 100 amperes per square foot. Nickel has a hydrogen overvoltage of approximately 1 volt which increases with current density.

Electrode materials in an electrolysis cell must be carefully selected to prevent corrosion and high overvoltages. Most soft metals have high overvoltages and corrode easily while the harder noble metals have low overvoltages and are less susceptible to corrosion.

In commercial electrolysis cells, nickel anodes and steel cathodes are used. The cathode is in a reducing atmosphere and therefore materials normally resistant to caustic may be used. However, since the anode is in an oxidizing atmosphere, nickel or a noble metal must be used to prevent the corrosive formation of oxides.

The Zero Gravity Problem

Electrolysis of water is a well established industrial process for the production of pure hydrogen and oxygen. As the gas bubbles are generated, gravitational forces separate them from the electrodes. The electrolyte is constantly in contact with the electrodes. Under zero gravity conditions the following problems are encountered:

a. The bubbles do not separate from the electrodes.

b. The electrolyte must be contained in a manner to prevent it from "floating away" from the electrodes.

c. Hydrogen and oxygen bubbles adhere to the electrodes and can cause complete polarization of the cell by preventing the electrolyte from contacting the electrodes.

In order to select the most feasible approach from which to attack these problems, a literature survey was planned. However, an acceptable literature survey was conducted by Battelle Memorial Institute as reported in MRL-TDR 62-44. In this survey, several methods of electrolyzing water were established. The reverse fuel cell type of electrolysis cell appeared to be the most promising for aerospace applications for the following reasons:

a. The mechanism for removal of evolved gases utilizes the least number of moving parts.

b. Standard electrolyte and electrode materials may be utilized.

c. Extensive research has been done on fuel cells.

The reverse fuel cell may be described as operating as follows: Water and electrolyte are contained in the absorbent medium between the wire screen electrodes independently of gravity. Close spacing between the electrodes eliminates the possibility of concentration polarization. The gases are evolved from the voids between the screens thereby eliminating the zero gravity problem of liquid-gas separation.

Experiments: Single Cell

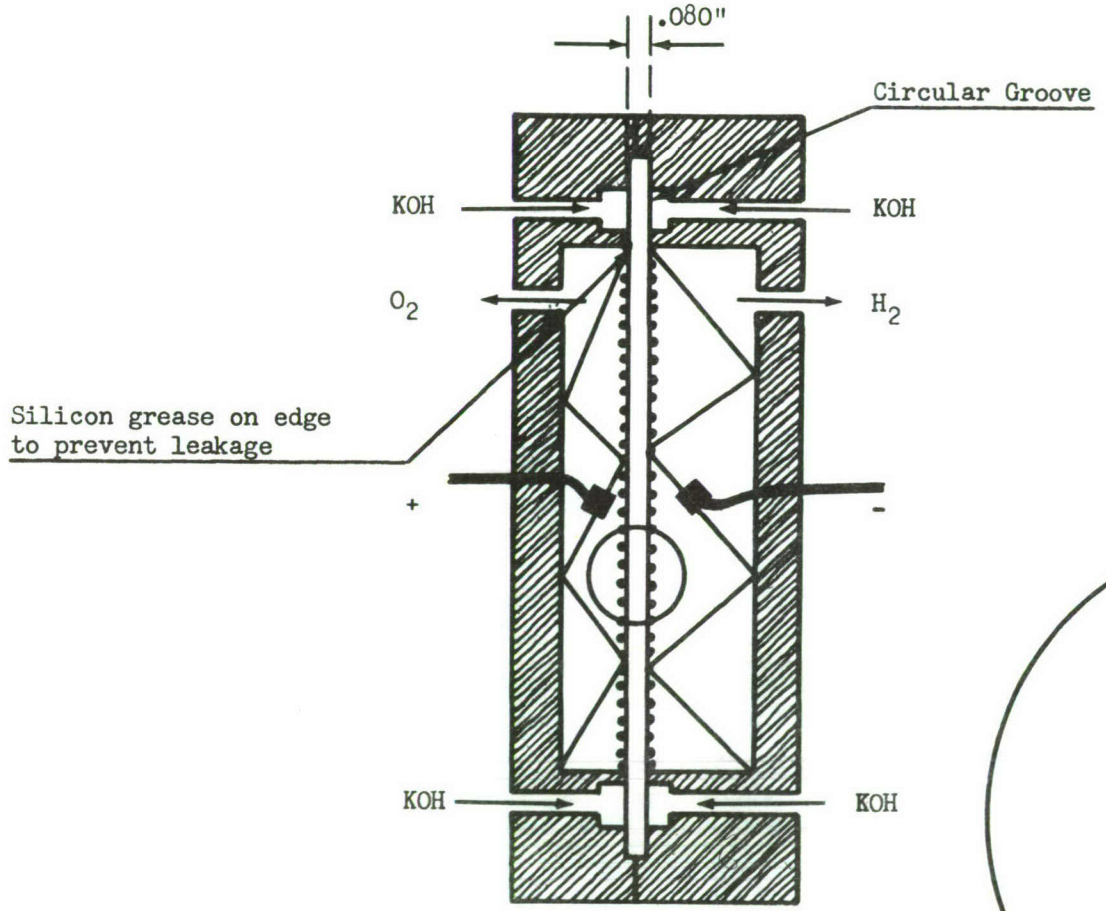
The initial experimental work done on this project was to determine the best materials for construction of a reverse fuel cell type of electrolysis cell. These materials include electrodes and an absorbent matrix. A single test cell was fabricated in the shops. (See Figure 1)

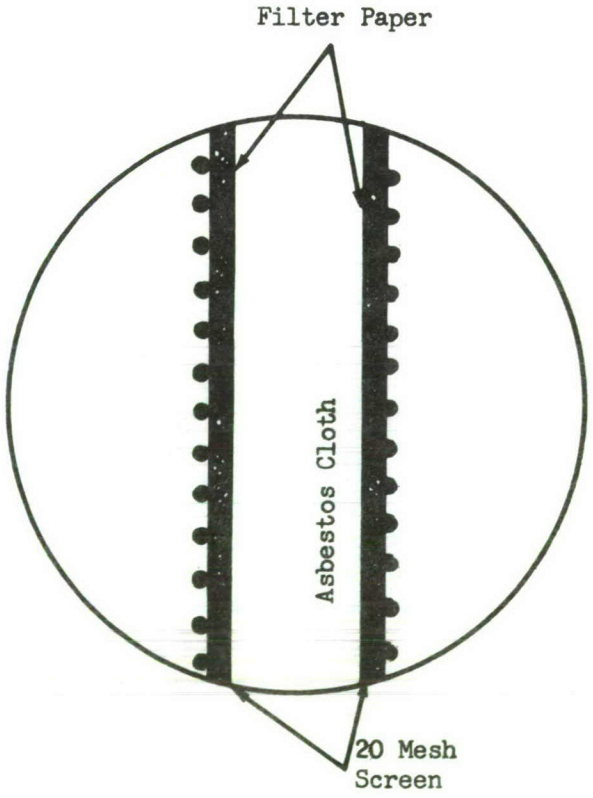
Initially, a design utilizing KOH supported in purified wood cellulose between asbestos sheets and nickel screen electrodes appeared feasible. Ordinary asbestos sheeting was soaked in 2 normal (N) KOH with no apparent effects. However, asbestos deteriorated when soaked in hydrochloric acid.

To determine the best design for an operating electrolysis cell, the following cells were fabricated and tested:

a. The initial cell of this series was fabricated with nickel screen (20 mesh)electrodes having a nominal area of 0.1 square foot. The absorbent matrix consisted of two 1/8 inch asbestos (laboratory grade) sheets on either side of 1/4 inch of absorbent cellulose saturated with

FIGURE I





ELECTROLYSIS CELL DESIGN

1 N KOH. The cell was wetted with the KOH, but ohmic resistance was high (30 ohms). It was concluded that the ion concentration was too low for good electrical conduction.

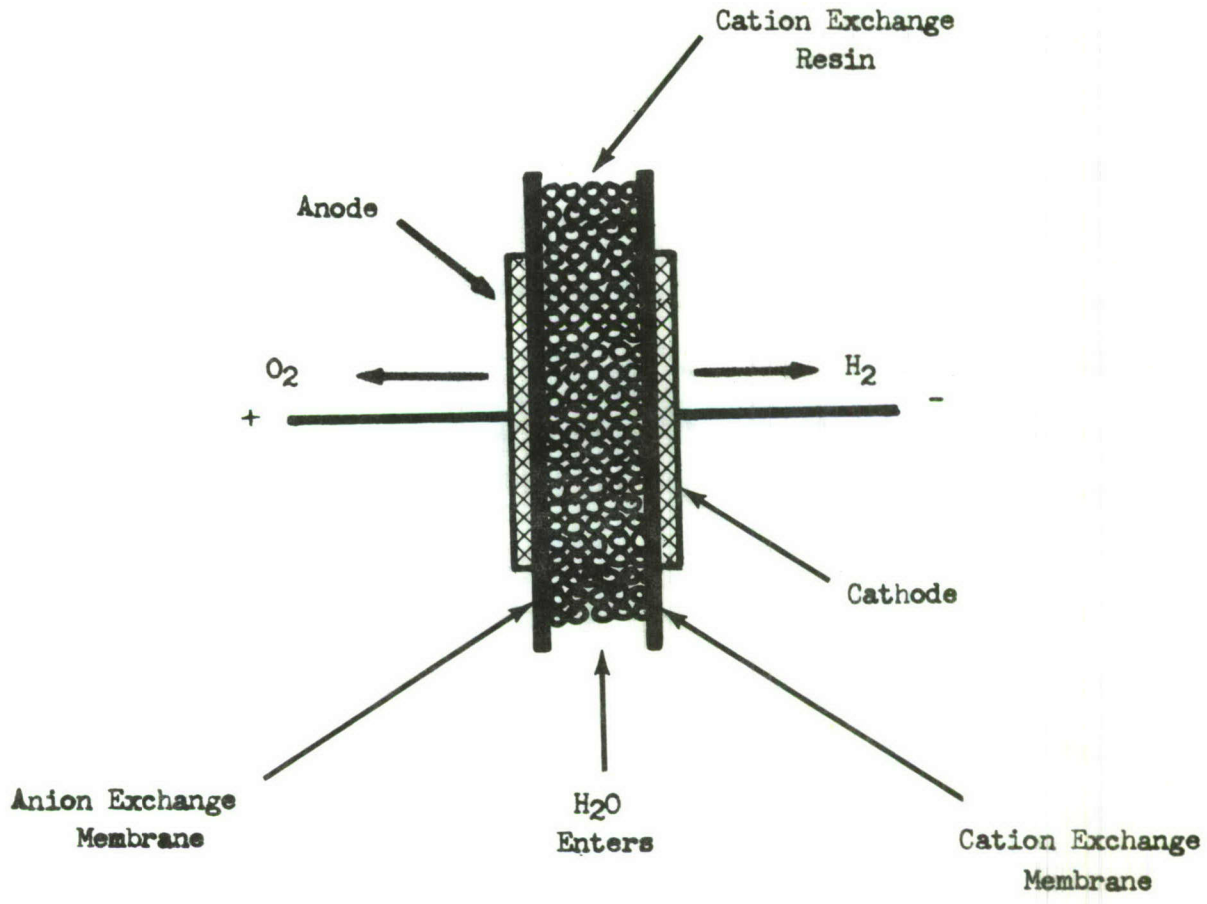
b. The same cell was taken apart and reassembled using orlon felt between the asbestos instead of cellulose; 5 N KOH was used as the electrolyte. The cell had slightly lower resistance but still no bubbles were apparent; only 2 amperes were passed by 12 volts. After the cell was disassembled, it was observed that the asbestos had lost body attaining a gelatenous-like character. The capillary structure had been destroyed and, as a result, ionic mobility was poor.

c. The failure of laboratory asbestos to work satisfactorily led to the investigation of a cell using pure water and cation exchange membranes. (See Figure 2) This type of cell operates as follows: When water enters the cation exchange resin, it is "split" into H^+ and OH^- ions. The H^+ ions are carried on the resin toward the cathode through the cation membrane where they take on electrons and becomes gaseous molecular hydrogen. The OH^- ions are carried through the anion membrane to the anode where the following reaction takes place:

$$4OH^- \longrightarrow O_2 + 2H_2O + 4e^-$$

A cell of this type was constructed using the nickel screen electrodes that were used in the previous test cells. It was found that no current was measureable below 5 volts and at 12 volts there was only 0.5 ampere. It appears that the net ion concentration from the water splitting was too low to be of practical significance.

FIGURE 2



ION EXCHANGE CELL

d. A cell was fashioned using anion exchange membranes, a cellulose center, nickel plated electrodes and 5 N KOH electrolyte. The current started somewhat high at 6 amperes and dropped to a low of 4 amperes at 3 volts. Apparently the cell had both high ohmic resistance and ionic polarization. At 3 volts the nickel anode corroded as was expected from the higher voltage. Gases were evolving from pools of electrolyte at the bottom of the gas chambers; there was poor current flow through the membranes where they were only dampened with electrolyte.

e. The cell was disassembled and the cellulose was replaced with anion exchange resin; 5 N KOH was still the electrolyte. The operation of this cell proved to be no better than the cellulose centered cell. In operating this cell it was observed that bubbles were evolving from the area under supporting rings. Apparently the membrane was not in 100% contact with the electrodes because of uneven support measure. This is a mechanical structure problem that can be solved in future cell designs. It was concluded that a wick on the membrane might possibly aid in holding electrolyte in contact with the electrode to provide better wetting and a more uniform current path.

f. The final cell of this series was assembled using a cellulose center (well packed) and two filter paper wicks pressed against the inner side of the membranes. At 6 volts, the cell passed 9.4 amperes. On the H_2 side a pool of electrolyte was formed and bubbled violently. There was no pool nor visible bubbles on the O_2 side, but O_2 was evolving as was demonstrated by exhausting it through water. The current dropped to 6 amperes; water was added and the current rose slightly. It was concluded that polarization had occurred which was disturbed by the addition of water.

The pools of electrolyte formed at the bottom of the gas chambers may possibly be the result of diffusion through the membrane, condensation from the inner surfaces of the cell and/or, on the O_2 side, a product of reaction. This water, if left to accumulate, might eventually flood the cell.

From these experiments, the following conclusions were reached:

a. Electrodes must be clean, corrosion resistant, and have low overvoltages. Nickel is the most economical metal which meets these requirements.

b. Ion exchange membranes have high ohmic resistances which can cause self-damaging heat. A material that holds electrolyte by capillary action and has low ohmic resistance is most desirable, but it must also be resistant to H_2-O_2 diffusion through the matrix. Asbestos cloth possesses these properties and for this reason, it is used in commercial electrolysis cells. It was planned to continue experimenting with this material.

c. It is possible that water produced in the anode reaction might flood the cell if not readily absorbed near the anode.

d. Cell polarization can occur if the distance between electrodes is too great. For this reason the distance should be as small as possible.

After the initial studies on the single cell gave rise to conclusions as to the nature of the most desirable materials to be used for H_2O electrolysis cells, a method of introducing electrolyte to the cell was designed and the single cell was so modified in the shops.

The single cell was designed to feed the electrolyte to an asbestos cloth membrane via a groove cut in the membrane support flange. The nickel wire screens were held against the cloth by corrugated nickel plates to which wires were attached to supply current to the electrodes. The nominal electrode area of this cell is 0.1 square foot.

The cell was charged with 5 N KOH and operated at 2.5 volts and 4.5 amperes with a resulting current density of 45 amperes per square foot. The following day, the cell was operated at 2.5 volts, but performance deteriorated to 3.8 amperes. O_2 and H_2 mixing was detected and it was concluded that even a small pressure differential ($\triangle P$) across the cloth will cause mixing. It was decided that the cloth would have to be modified to allow a small $\triangle P$ (2 inches of water) across the cell without mixing.

The cell was assembled with low porosity filter paper on either side of the asbestos cloth. The cell was charged with KOH and operated at greatly reduced capacity. It was concluded that this filter paper hindered ion transferrence.

The filter paper was replaced with high porosity filter paper which is soft-textured laboratory grade. The cloth-filter paper combination was mounted with silicone grease on the support lip to prevent electrolyte leak-over and the electrodes, which were slightly corroded, were cleaned with HCl. The cell was successfully operated for seven hours at 2.5 volts at from 26 to 58 amperes per square foot with no hydrogen-mixing at pressure differentials in excess of 12 inches of water.

The electrolyte was added under constant pressure from a funnel head of approximately 18 inches of water.

This cell gave conclusive evidence that the asbestos cloth-soft filter paper membrane would provide good separation and current densities while wetted with electrolyte. The design of a six cell module was based upon this cell concept.

Experiments - 6 cell unit

Design of six cell unit:

General Considerations: The six cell electrolysis unit was limited in size to a cylindrical volume approximately 30 inches in diameter by 28 inches long with a gross weight of 300 pounds. The operating voltage in the Atlas Passenger Pod would be 28 volts supplied by a battery.

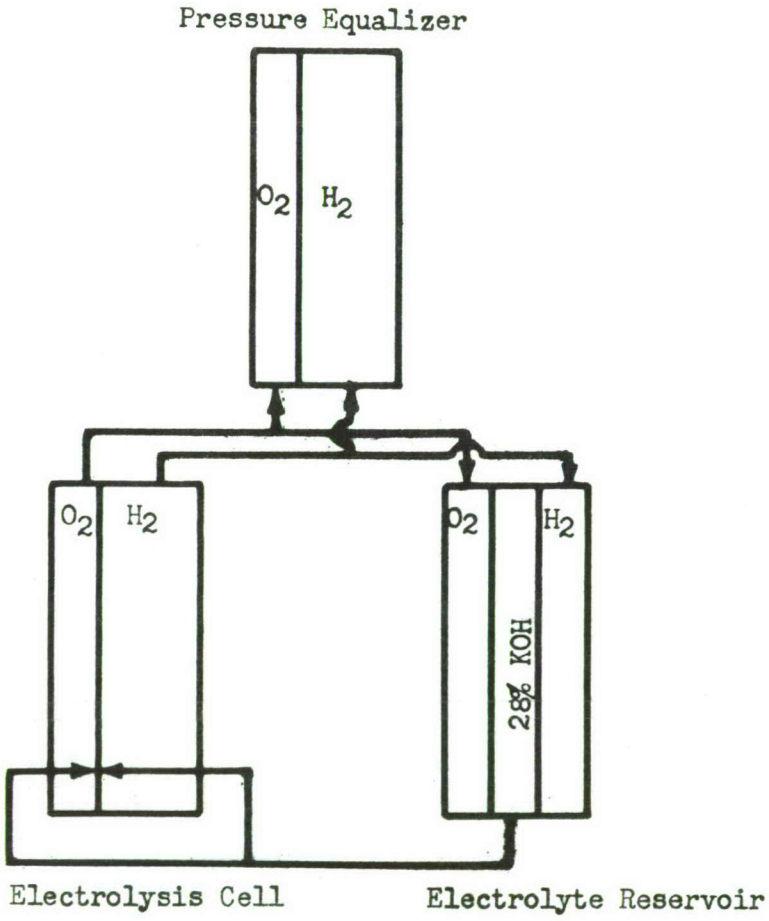
Design Considerations: The six cell unit was decided upon rather than 12 cells in order to simplify construction, cut construction time, and to economize on material requirements. The six cells would operate on 14 volts (a special 14 volt battery or a re-routed 28 volt battery) at a current density of approximately 40 to 50 amperes per square foot. The membrane size was .230 square foot which would require approximately 10 amperes. This power, then, would be compatible with a moderately sized battery on the Passenger Pod. At this current, the gas output of the cell is sufficiently low to avoid critical outlet vent sizing and high rates of liquid feed to the cell.

The final design of the cell would require a system capable of feeding water to the cell in zero gravity. A pressurized liquid feed system would work satisfactorily as would a feed system of wicks. Since both of these methods have been experimentally tested, the self-contained third method of having an internal pressure feed appeared to present an alternate unproven choice for a zero-gravity experiment. The pressure feed system designed for the cell operates as follows. (See Figure 3). Pressure from the cell is applied to each side of the electrolyte in the reservoir through flexible diaphragms (See Figure 4). As water is electrolyzed, the asbestos cloth begins to dehydrate and the capillary action demands more electrolyte. This action literally draws electrolyte from the reservoir thereby creating a void which is immediately filled due to gas pressure on the flexible reservoir. Elimination of voids in the electrolysis feed system insures that there will be a continuous stream of electrolyte from the reservoir to the electrolysis cell under weightless conditions.

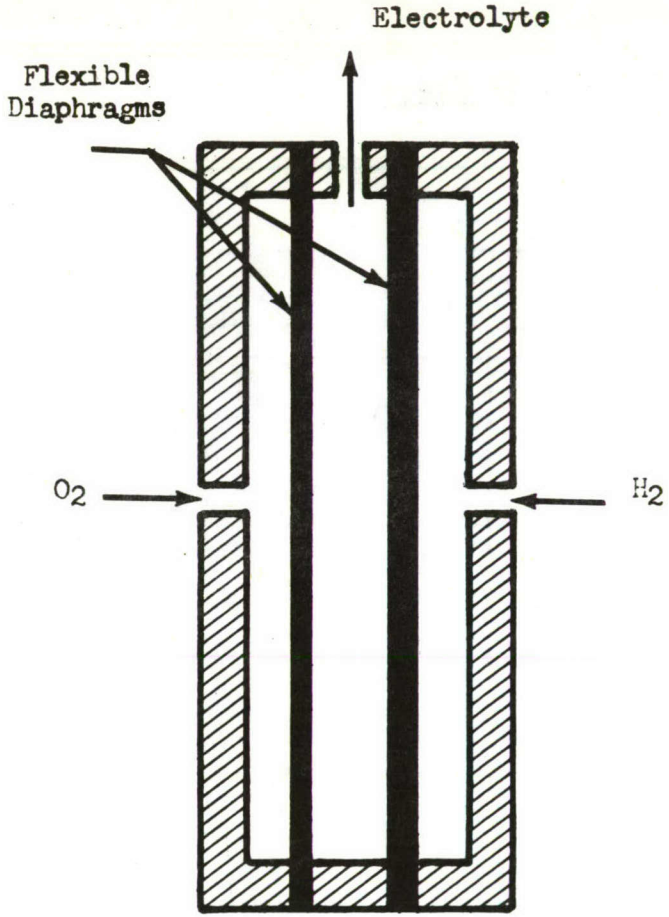
Since the asbestos cloth-filter paper membrane can withstand only slight pressure differentials, a pressure equalizer was incorporated with the cell (See Figure 5). It is a chamber having a 2:1 volume ratio with a flexible diaphragm separating the O_2 and H_2 compartments to keep the pressure-volume products constant.

The six cell unit then, is a cylindrical bi-polar electrolysis cell with the electrolyte reservoir at the forward end, and the pressure equalizer at the aft end. The construction of the cell is illustrated by Figures 6, 7, 8, and 9.

FIGURE 3

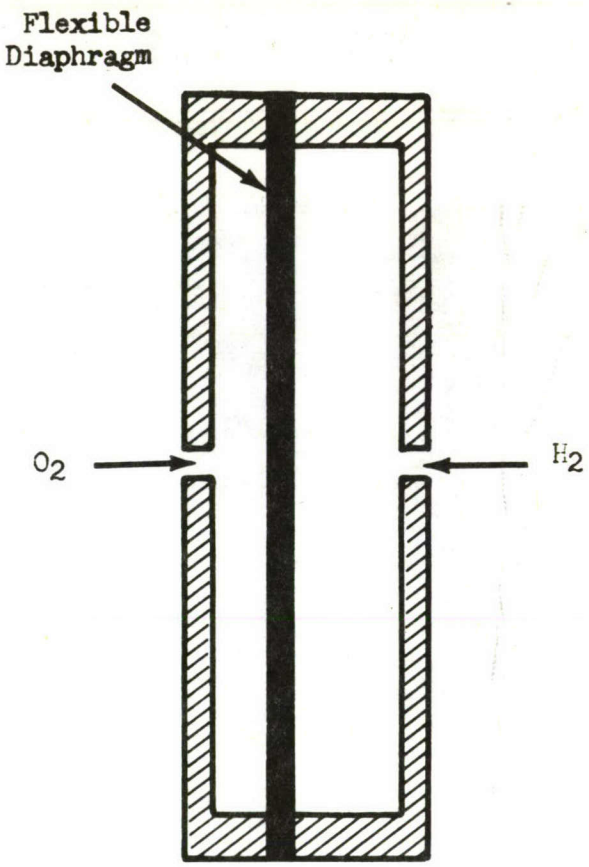


PRESSURIZED FEED SYSTEM FOR SIX CELL UNIT



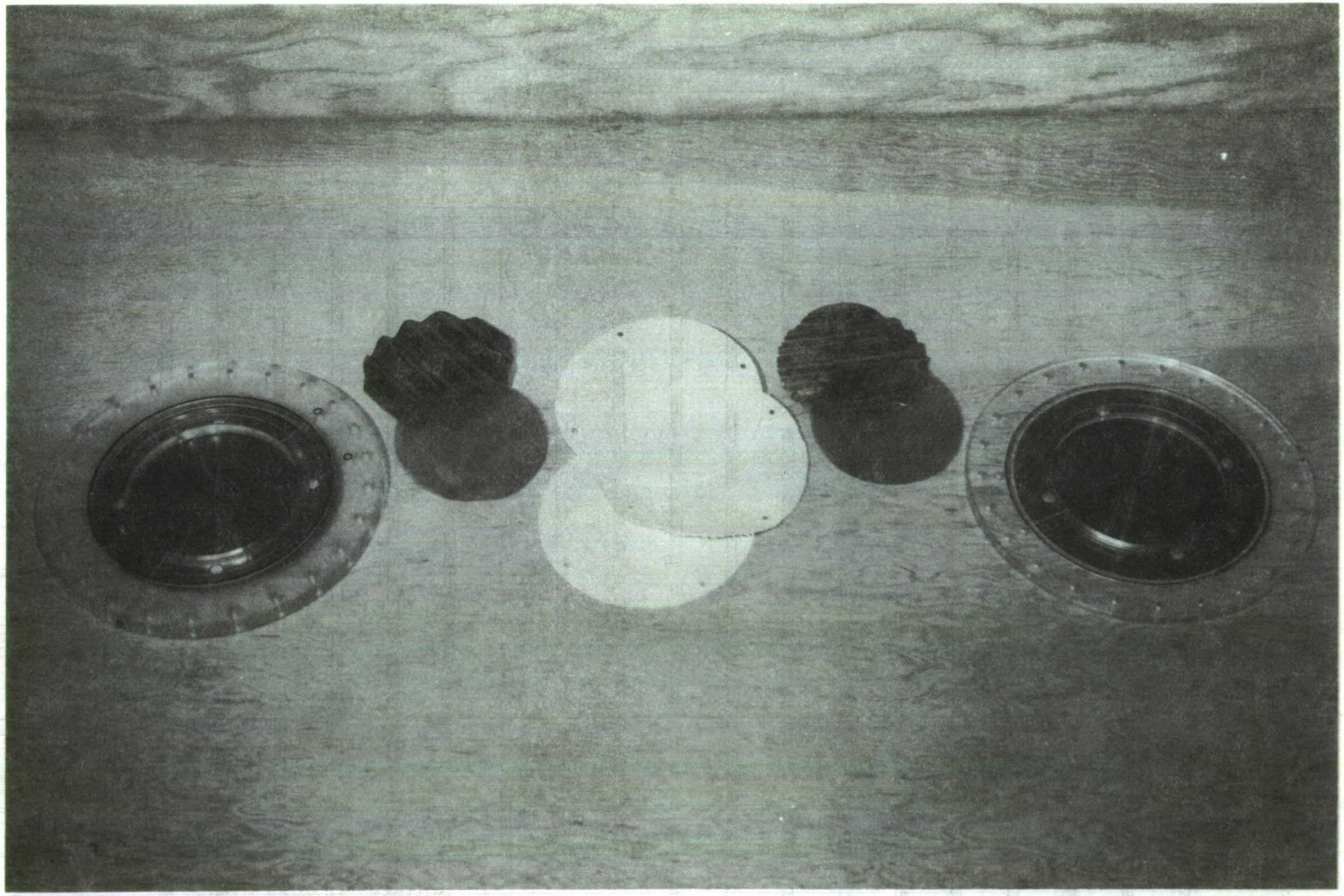
ELECTROLYTE RESERVOIR

FIGURE 4



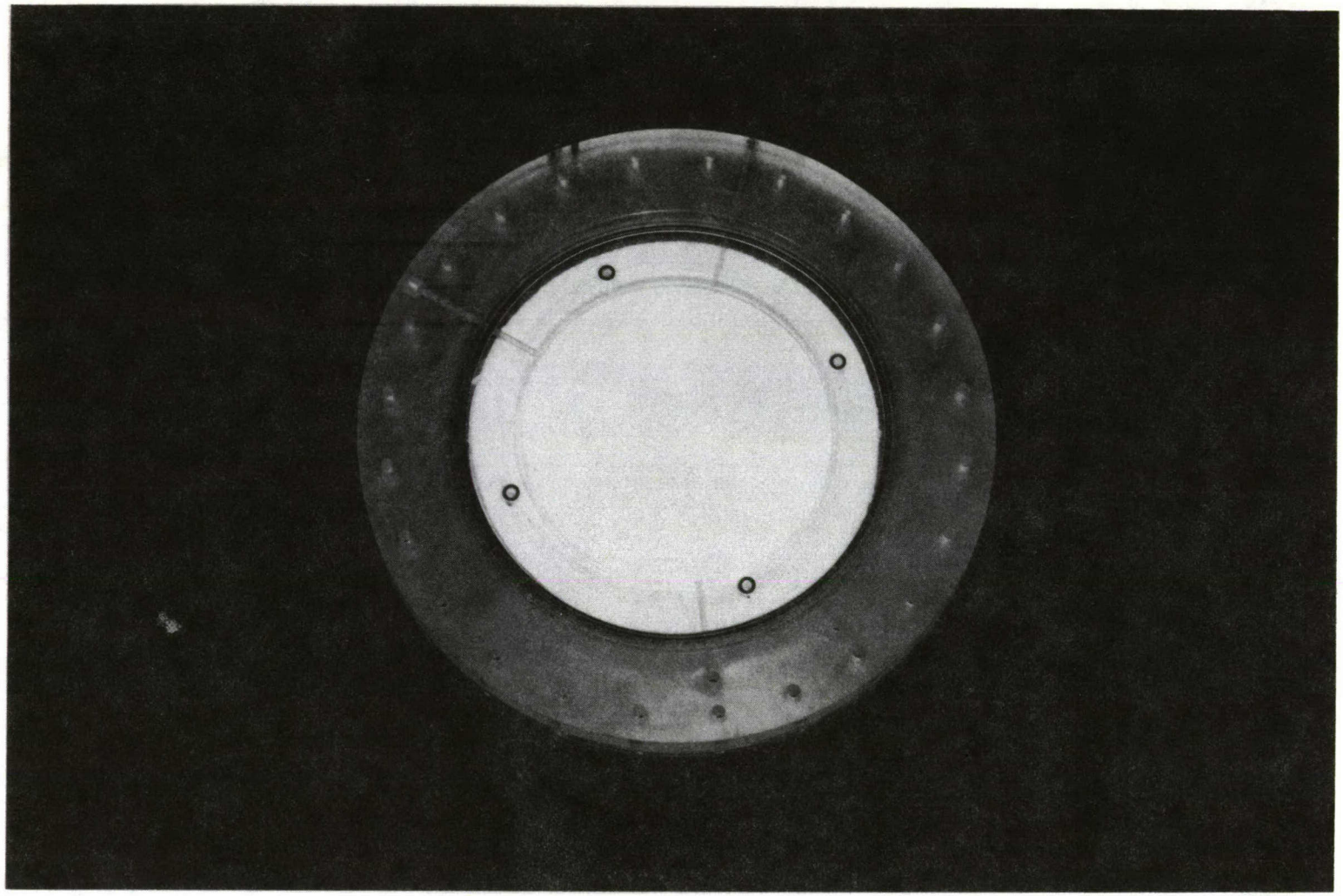
PRESSURE EQUALIZER

FIGURE 5



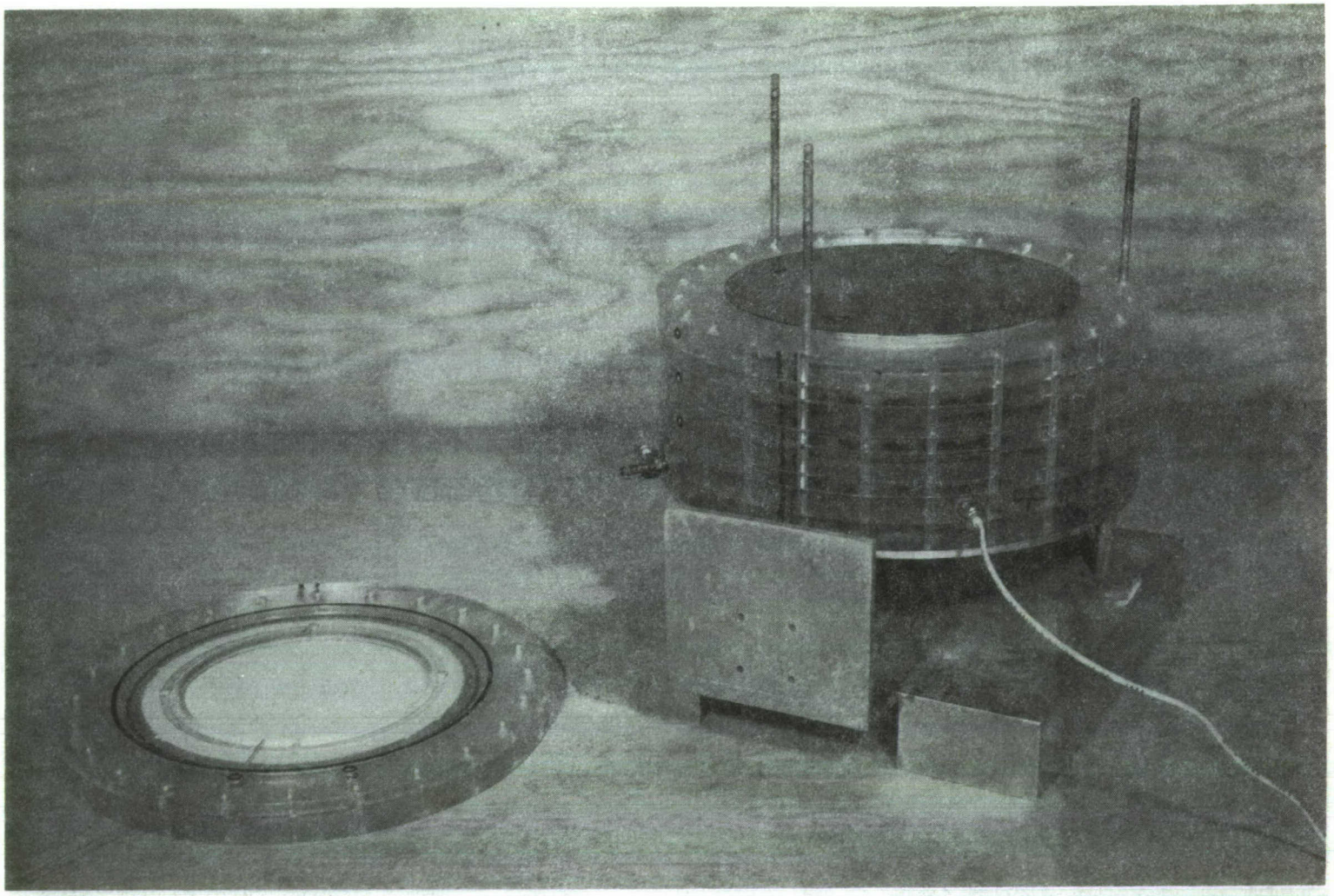
ELECTROLYSIS CELL
INDIVIDUAL CELL COMPONENTS

FIGURE 6



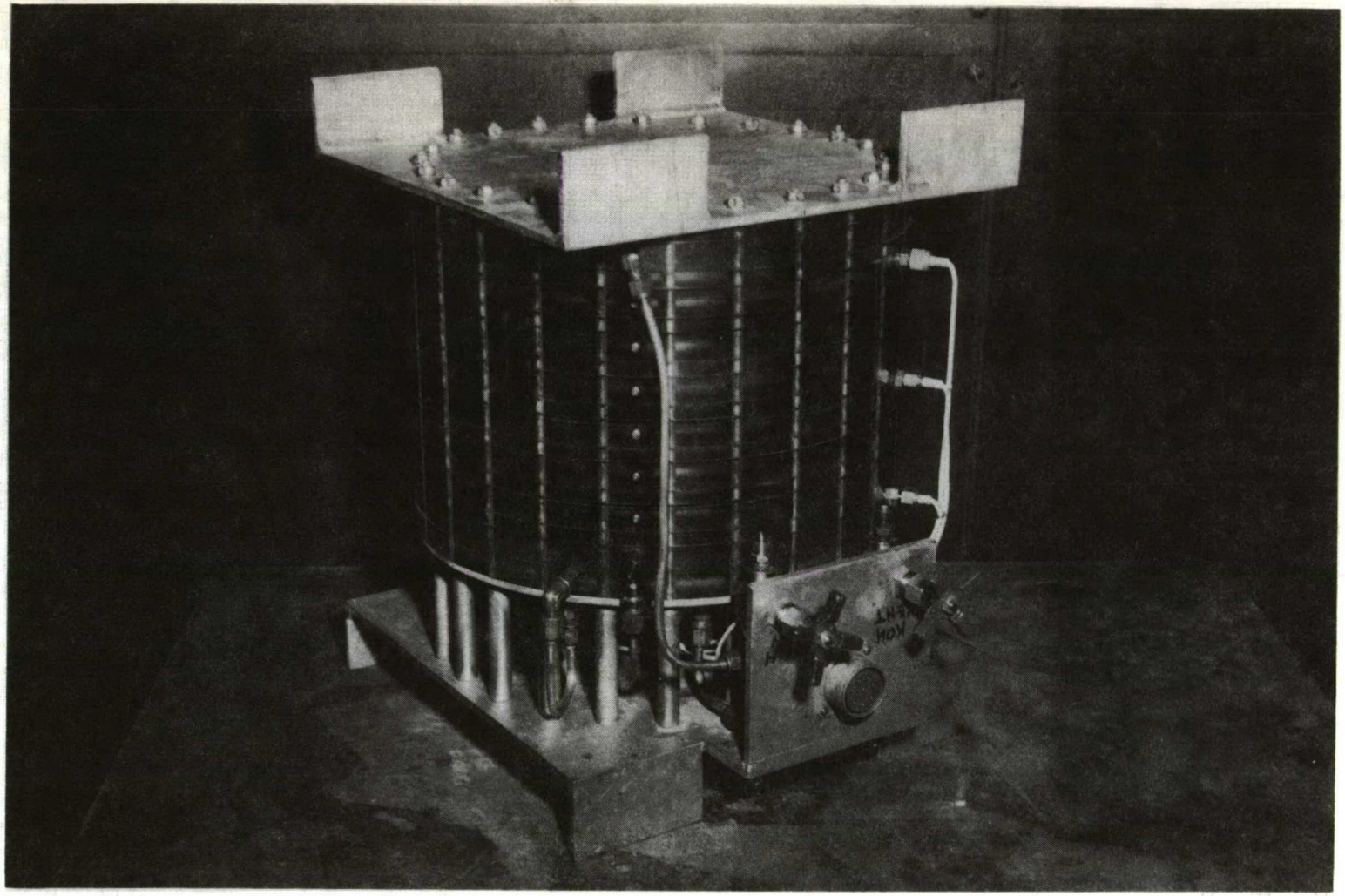
ELECTROLYSIS CELL
INDIVIDUAL CELL ASSEMBLED

FIGURE 7



ELECTROLYSIS CELL
PARTIAL ASSEMBLY OF POD TEST MODEL

FIGURE 8



"ELECTROLYSIS CELL"
PRIOR TO POD INSTALLATION

FIGURE 9

The six cell unit was assembled and laboratory tested. Figure 10 shows a typical performance curve while operating at 14 volts. The design introduced several problems which include the following:

a. The small "O" ring seals in the water manifold tended to leak because of insufficient seating by shifting position. The problem was overcome by enlarging the holes in the bipolar plates. The problem was also overcome by using small pieces of polyvinylchloride in place of the "O" rings.

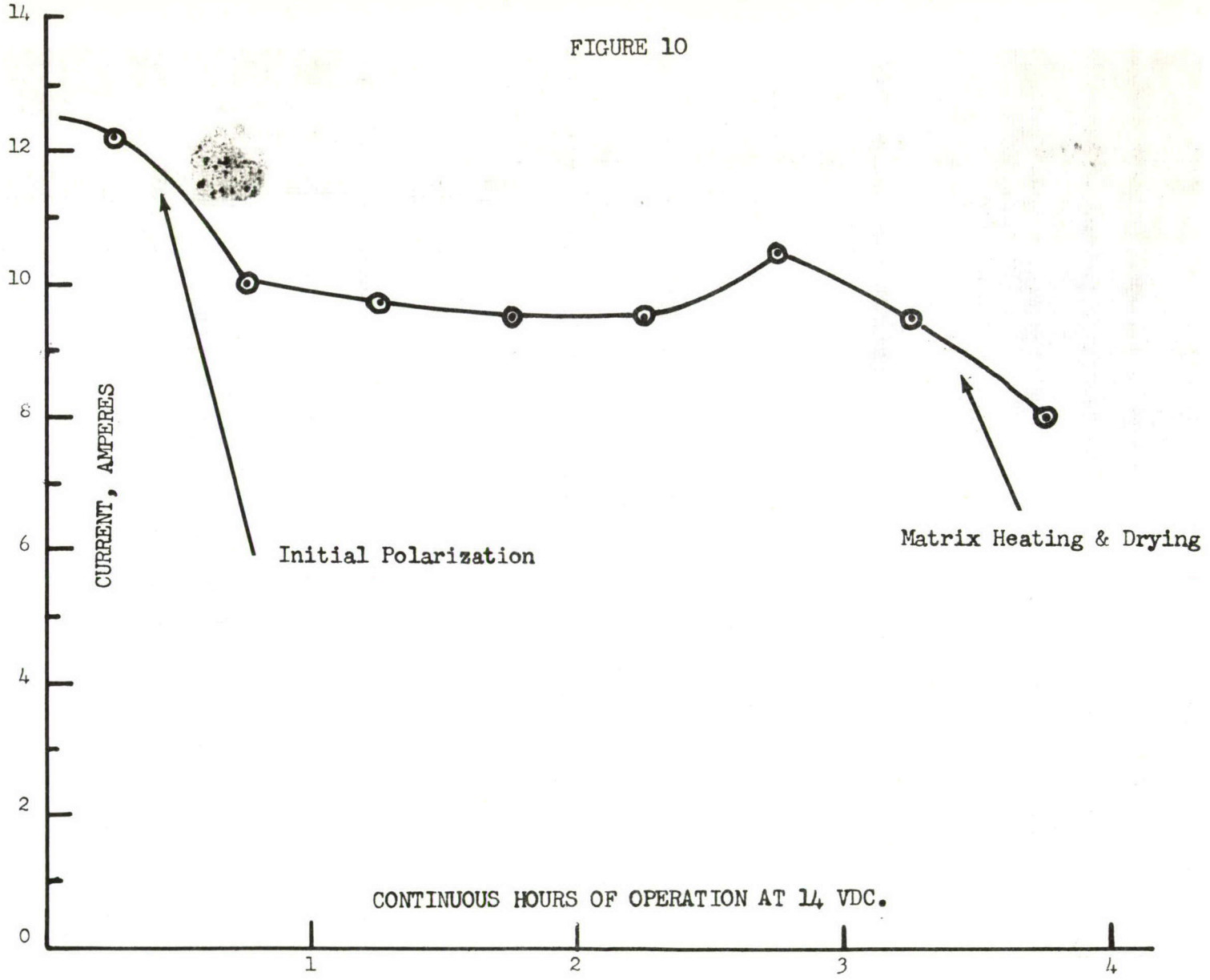
b. Minor leakage was encountered through the stranding of the wire leading into the cell. This problem was solved by sealing the terminal end of the wire which was inside the cell.

c. After several hours of cell operation at high current densities, the cell became warm and evidence of matrix drying was observed by slight hydrogen-oxygen mixing and decreased performance. The matrix, although a good liquid absorber, loses its capillary properties under the influence of higher temperature.

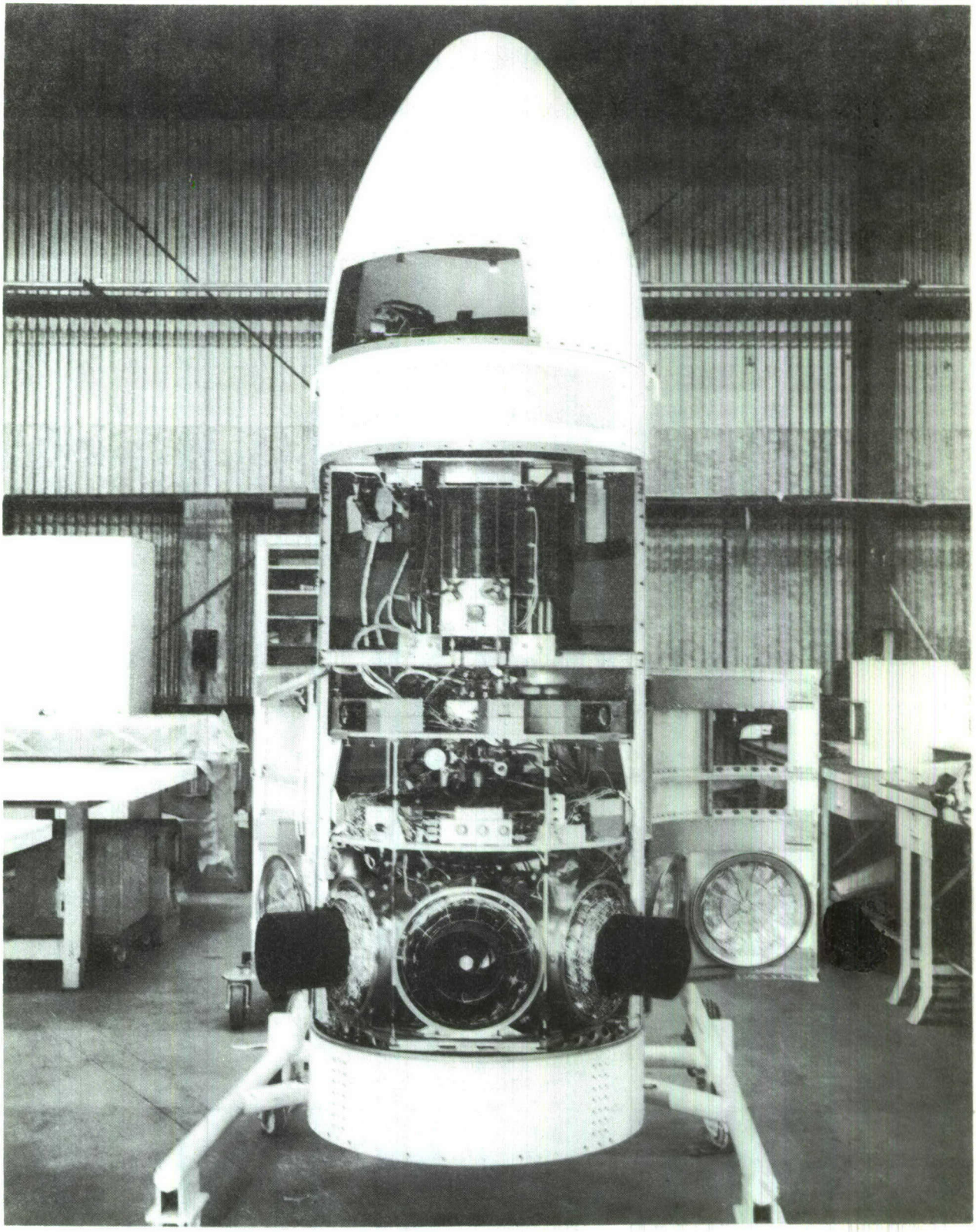
d. Foaming of electrolyte at high current densities is the result of liquid electrolyte being carried over by the gas formed at the electrode - matrix interface. This problem is inherent with high current densities in this cell design.

Figure 11 shows the electrolysis cell subsequent to integration into the Atlas Scientific Passenger Pod at General Dynamics/Astronautics, San Diego, California.

FIGURE 10



PERFORMANCE CURVE



ELECTROLYSIS CELL AFTER INSTALLATION IN ATLAS SCIENTIFIC PASSENGER POD

FIGURE 11

Conclusion:

Other than minor design deficiencies, the critical component of the electrolysis cell appears to be the matrix which separates the electrodes. This matrix must not only be capable of readily absorbing the electrolyte while resisting hydrogen-oxygen mixing through diffusion but must also permit electrochemical conductance with low I^2R heat production. Future in-house studies are planned directed toward the development and/or procurement of improved matrices capable of meeting these specifications.

One possible solution to the problem of electrolyte contamination of the evolved gases due to foaming would make use of semi-permeable membrane coverings at the exit gas ports in the cell interior. These hydrophobic membranes would permit the free diffusion of gases but prohibit the passage of electrolyte.

Although additional work must be done before the problem of developing a reliable water electrolysis system for use in weightless environments is completely solved, this project has brought to the surface some of the problems commonly associated with electrolysis cells that, to the casual observer, on paper appear free of problems. Thus, this program has helped develop an Air Force proficiency in both designing and evaluating electrochemical atmosphere control processes.

(U) INVESTIGATION OF SEMICONDUCTOR EQUALIZER DIODES FOR SECONDARY BATTERY CELLS

by

Wayne Scott Bishop
Civilian
GS-11

Air Force Aero Propulsion Laboratory (APIP-2)
Wright-Patterson Air Force Base, Ohio

INVESTIGATION OF SEMICONDUCTOR EQUALIZER DIODES FOR SECONDARY BATTERY CELLS

Wayne Scott Bishop

SUMMARY

This paper considers the problem of sealed nickel-cadmium cell unbalance and the effect of this cell unbalance on battery performance. Methods for controlling cell unbalance and improving battery performance are given. A method employing semiconductor diodes for the control of each individual cell in the series string is considered in detail. Results of experiments with nickel-cadmium cells cycled continuously at deep discharge and rapid charging to simulate low earth orbit conditions are presented.

INTRODUCTION

This paper considers the problem of cell unbalance in secondary nickel-cadmium batteries and methods for its control.

Due to a great many variables which are virtually uncontrollable during the manufacture and use of sealed nickel-cadmium cells, they tend to become unbalanced when cycled in a series connected string [1]. This unbalance is due principally to differences in charging efficiency and is characterized by two phenomena called "overcharge" and "reverse charging". Since cells are used as a series connected group or battery they are charged from a common power source. Some cells which have higher charging efficiencies attain full charge before other cells. If current is continued to be passed through these cells with higher charging efficiencies until all cells have attained a full charge, those with the higher charging efficiency are said to be "overcharging". During overcharging, no additional energy is

stored, electrical energy is converted to thermal energy which results in reduction of cell life. Reverse charging results when one or more cells in a series connected string have liberated all of their stored energy and yet a relatively large number of the remaining cells are capable of delivering energy to the load. The cells still capable of delivering energy will force current through the depleted cells and charge them in a reverse direction. Both overcharging and reverse charging shorten battery life. These problems are further amplified by deep discharges and rapid charging [1, 2].

To achieve long life, current satellite batteries use only 10 to 25 per cent of their capacity ratings. The unbalance problem is not as severe in this regime; however, the battery specific energy (W-H/#) is significantly reduced. Currently, battery weight averages about 10 per cent of the total satellite weight [3]. If cell unbalance can be controlled and capacity utilizations increased to about 75 per cent for long cycle lives, the battery weight could be reduced to about 3 per cent of the total satellite weight.

There are three methods being investigated for control of cell unbalance. These three methods are: coulometers, auxiliary electrodes, and semiconductor diodes. The coulometer and auxiliary electrode methods of control will be briefly discussed later in this paper. The diode method will be discussed in detail and this discussion will include: diode theory and its application to cell balancing; description of the charge control diode; the interaction of the cell and equalizer circuit, and experimental results for a three cell nickel-cadmium battery with the

diode equalizer circuit. The nickel-cadmium battery will be considered since it has been the most widely used and is the only battery being considered for long life missions such as communication satellites.

TECHNICAL DISCUSSION

The Cell

The problem of cell unbalance [1] exists in all secondary alkaline battery couples. The nickel-cadmium battery is probably the most familiar of the alkaline couples since it has been widely used for industrial and aircraft applications, as well as numerous products for the home such as electric toothbrushes, electric razors, and electric drills. The nickel-cadmium battery exhibits a charging efficiency in the range from 77 to 91 per cent [2] depending upon such things as the type of charging used (constant current or constant voltage), the charge time, the temperature, and many other variables. A major factor which results in this widely varying charge efficiency is the use of a porous electrode. The porous electrode is a highly complex structure about which little is known. Due to formation of different compounds during charge and discharge the electrode structure changes, and it is virtually impossible to predict the exact behavior of a porous electrode from cycle to cycle. It is these factors which appear to account for cell unbalance. Due to variations in charging efficiency, some cells may reach full charge while some cells are only 80 per cent charged. In order to charge all cells fully, it is necessary to overcharge the higher efficiency cells.

During the overcharge period in a nickel-cadmium battery, oxygen gas and water are liberated at the nickel electrode according to the following reaction:

$$4\ OH^- \rightarrow 2\ H_2O + O_2 + 4\ e^- \qquad (1)$$

and four electrons are given up. Two hydroxyl ions are consumed from the electrolyte. The oxygen then diffuses to the cadmium electrode where it reacts with water and cadmium to form cadmium hydroxide by the following reaction.

$$O_2 + 2H_2O + 2\ Cd \longrightarrow 2\ Cd\ (OH)_2 \qquad (2)$$

The cadmium hydroxide is then charged back to cadmium metal and hydroxyl ions are released to the electrolyte according to the following reaction.

$$2\ Cd\ (OH)_2 + 4\ e^- \longrightarrow 2\ Cd + 4\ OH^- \qquad (3)$$

These reactions can continue for long periods of time since no reactants are consumed and they require only the addition of electrical energy. From a practical viewpoint the overcharge currents are usually limited to a maximum value of $C/10$ amperes, where "C" is the ampere hour rating of the battery. This is necessary since the rate at which the reactions occur depend upon oxygen partial pressure and temperature. Charging rates higher than $C/10$ result in excessive internal cell pressure. These higher overcharge rates can also result in voltages sufficiently high to hydrolyze the water in the electrolyte by the following reaction.

$$2\ H_2O \longrightarrow 2\ H_2 + O_2 \qquad (4)$$

The oxygen produced by hydrolysis can be recombined as previously described; however, the hydrogen will not readily recombine and will accumulate in the cell. This results in pressure rise within the cell and ultimately failure due to case rupture. The presence of oxygen can result in separator degradation due to oxidation and ultimately cell failure.

The maximum voltage at which the cell can be charged is dependent upon the time which the cell is charged and the temperature. Typical maximum voltages for nickel-cadmium cells at room temperature range from 1.55 volts per cell for short charge periods (one hour) to 1.45 volts per cell for continuous charging (float). Since the cells are used as series connected strings, the charging voltage for the string is controlled. This method of charging does not provide direct control of individual cell charging voltages.

During reverse charging, hydrogen is again evolved. To prevent this hydrogen evolution the cell potential should always be greater than zero; however, this is difficult to achieve. Experimental data has shown [4] that it may be possible to permit cell polarity reversal if the voltage is maintained below minus 0.5 volt.

The methods of charge control described below may be employed to provide cell protection.

The Coulometer

There are several types of coulometers, such as electronic integrators, moving fluid in a capillary tube, and electrochemical cells with like

electrodes. These devices all protect the battery rather than individual cells. They are sized so that when the design capacity of the battery has been removed, an electric signal is provided. This signal can be used to remove the battery from the circuit. Also, during charging, the coulometer will provide an electric signal when the capacity delivered to the battery is equal to the net amount removed from the battery during previous cycles. This signal can be used to remove the battery from charge or reduce the current to a safe value for overcharging. The coulometer provides some degree of overdischarge protection for the battery; however, no provisions for cell reversal control can be made. It has the advantage of being simple and relatively inexpensive; however, it would seem possible that it can suffer many of the same difficulties of the battery cells such as separator oxidation, seal leakage, and electrode growth.

The Auxiliary Electrode

This method of controlling cell unbalance consists of placing a catalytic or active surface in the cell where it can easily promote oxygen recombination. This electrode can be used to recombine oxygen or to provide a voltage signal indicating the presence of oxygen. There are two different types of these electrodes, one which operates similar to the cadmium battery electrode in the previously described recombination mode. The other type electrode behaves similar to the oxygen fuel cell electrode according to the following reaction.

$$2\,H_2O + O_2 + 4\,e^- \longrightarrow 4\,OH^- \qquad (5)$$

These electrodes can be used for recombination of oxygen at low and

moderate charging currents. At high charging currents, the current density is too high and results in poor recombination performance of the auxiliary electrode; however, these electrodes can be used to provide an electrical signal. This signal can be used to remove the cell from charge, remove the battery from charge, or reduce the charging current to a safe overcharge level. If this electrode is used for recombination, external resistors are required. If it is used for signaling the presence of oxygen, scanning and switching mechanisms are required.

This method has the advantages of individual cell control and the ability to sense the presence of oxygen. Its disadvantages are the requirement for external electronic circuitry, no cell reversal protection, and the requirement for an additional hermetic terminal seal. Hermetic terminal seals are currently one of the major problems in achieving a long life communications satellite battery.

The Diode Control

The semiconductor equalizer method for controlling cell unbalance provides control of each individual cell voltage during overcharge and overdischarge. The semiconductor control method employs two diodes connected in parallel with each cell as shown in Figure 1. The diode called the "Stabistor" provides control of the end-of-charge voltage and it has a volt-ampere characteristic as shown in Figure 2. Ideally, the diode would appear as an open circuit below its threshold voltage (1.48 - 1.50 volts). Above this voltage, it would appear as an ideal shunting circuit. This would permit each cell to accept charge until it

reached a maximum safe charging voltage (in this case 1.50 volts). Above the maximum voltage, the diode would conduct and shunt all of the current around the cell. The characteristic for the actual diode indicates that approximately 1 ampere is shunted around the cell at a charging voltage of 1.50 volts, and that the slope of the curve increases sharply above this voltage. Figure 3 gives a typical unprotected nickel-cadmium cell terminal voltage characteristic for a constant current charge. Figure 4 shows a typical terminal voltage characteristic for a diode protected nickel-cadmium cell with constant current charging. The diode current is also shown in Figure 4.

The other diode shown in Figures 1 and 5 labeled "Diode for Reverse Charging Control" ideally acts as an open circuit until the cell voltage reaches zero, at which time it begins to conduct and shunt the discharge current around the cell. It has not been possible to achieve this in actual diodes, but cell voltages can be maintained below -0.5 volt. Figure 6 shows a volt-ampere characteristic for an actual and ideal cell reversal control diode. The maximum value of negative voltage to which a cell can be permitted to reverse has not been established but has been dictated by the diode. Additional research is required to establish the negative voltage level to which a cell may be reverse charged without deleterious effects.

Special germanium diodes were fabricated for reversal protection; however, P-N-P germanium power transistors connected as shown in Figure 7 were found to perform as well as the special diodes and are less expensive.

The "Stabistor" diode is a two junction silicon type. It utilizes a 5 milliohm-centimeter, P-type silicon starting material. Phosphorous is diffused into the silicon to form the N-type region. Two of these elements are then tin soldered in series to form the active portion of the device.

In the design of a battery system employing diode control, the diode and cell characteristics and their interactions must be considered. Among these interactions are: sizing of the diodes to the current requirements; cell voltage variation with temperature; heat sinking of the diodes; fully charged stand time; thermal coupling of the diode and cell; mechanical problems; reliability; and length of the charge period. The Stabistor threshold voltage must be selected in accordance with the charge time available. An end-of-charge voltage limit at room temperature of 1.55 volts per cell is permissible if short repetitive cycles such as 90 minutes, are used [2]; however, if long charge times (10 hours or longer) are used, the end-of-charge voltage limit should be reduced to about 1.48 volts/cell. The magnitude of the charge and discharge currents must be determined in order to fix the diode power ratings. The end-of-charge voltage has a negative temperature coefficient and is somewhat dependent upon the charging rate [2]. The "Stabistor" must be designed so that its threshold voltage duplicates (as nearly as possible), at each temperature, the desired end-of-charge voltages. The power dissipation of the diodes must be considered in the choice of heat sinks and thermal coupling.

Figure 8 shows a 3-cell, 3.4 ampere-hour experimental nickel cadmium battery with diode equalization. The diodes and cell shown in Figure 9

were mounted in a one-fourth inch copper block to aid in thermal coupling. The electrical schematic with paths of overcharge and overdischarge currents labeled is shown in Figure 10. This battery was cycled for 2,764 times at room temperature and a discharge equal to 75 per cent of the rated capacity or 2.6 ampere hours [5]. Each cycle consisted of a constant current charge at 3.4 amperes for 55 minutes and a constant current discharge at 4.5 amperes for 35 minutes. Similar batteries that were unprotected failed after 250 cycles under identical conditions [2]. Originally, it was planned to cycle two unprotected cells in series with the protected cells; however, they had to be removed after 251 cycles due to excessively high end-of-charge voltages (1.6 - 1.7 volts). There was a danger of cell case rupture since excess gas pressure could result from these high voltages. Figure 11 shows a curve of cell end-of-charge voltage versus cycle number for both the protected and unprotected cells [5 & 6].

During open circuit stand the "Stabistor" diode will discharge the cell by allowing leakage current to pass through it [4,5]. This leakage current results because the stabistor diode conducts current at a voltage below the open circuit voltage of the cell. Leakage currents as high as C/20 have been measured.

The main problems still to be investigated are those of heat sinking and thermal coupling, reduction of diode leakage current, and determination of the exact cell reversal process and its effects.

Current efforts include design, fabrication and test of experimental prototype cells with the diode as an integral part; reduction of the diode

leaking current; and adaptation of this control system to the silver-cadmium battery.

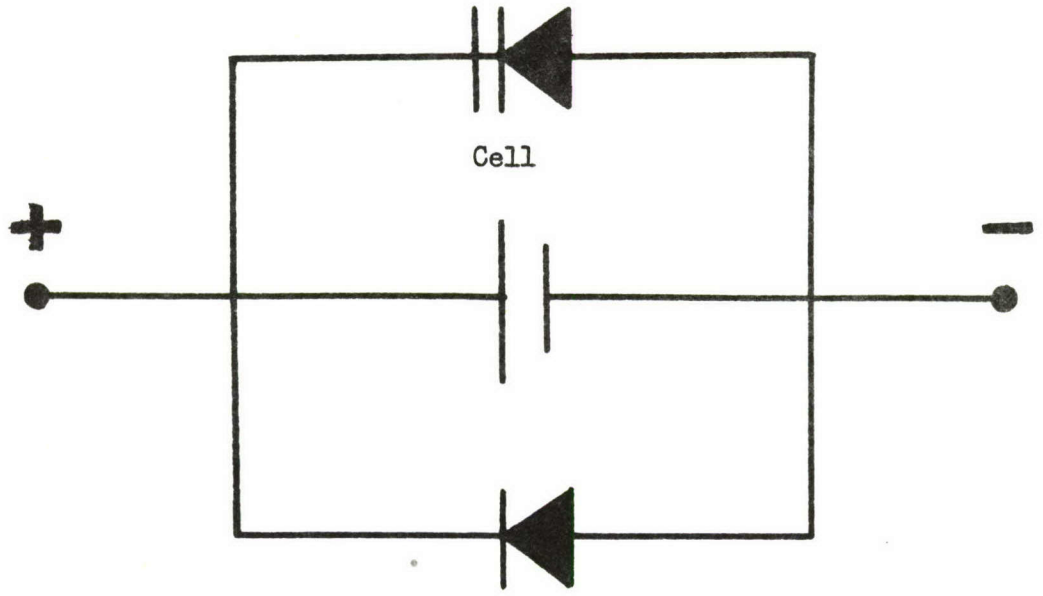
CONCLUSIONS

The diode control method offers promise of effectively improving nickel cadmium battery specific energy and/or cycle life.

Future work should include a more detailed investigation of cell reverse charging to determine its effects and establish safe voltage limits.

D I O D E C O N T R O L C I R C U I T

Stabistor



Diode for Reverse Charging Control

Figure 1

VOLT-AMPERE CHARACTERISTIC FOR AN ACTUAL AND IDEAL STABISTOR

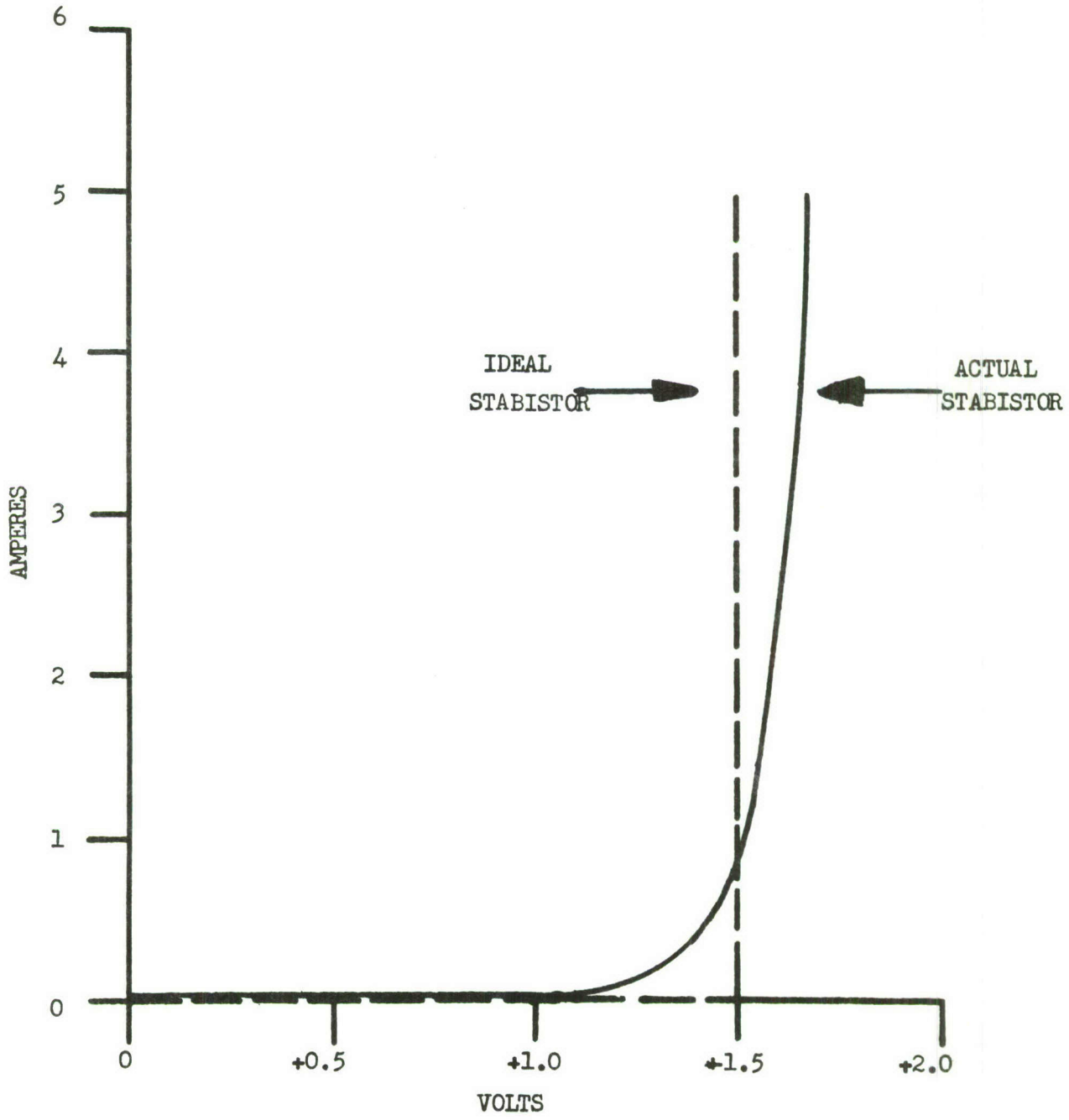
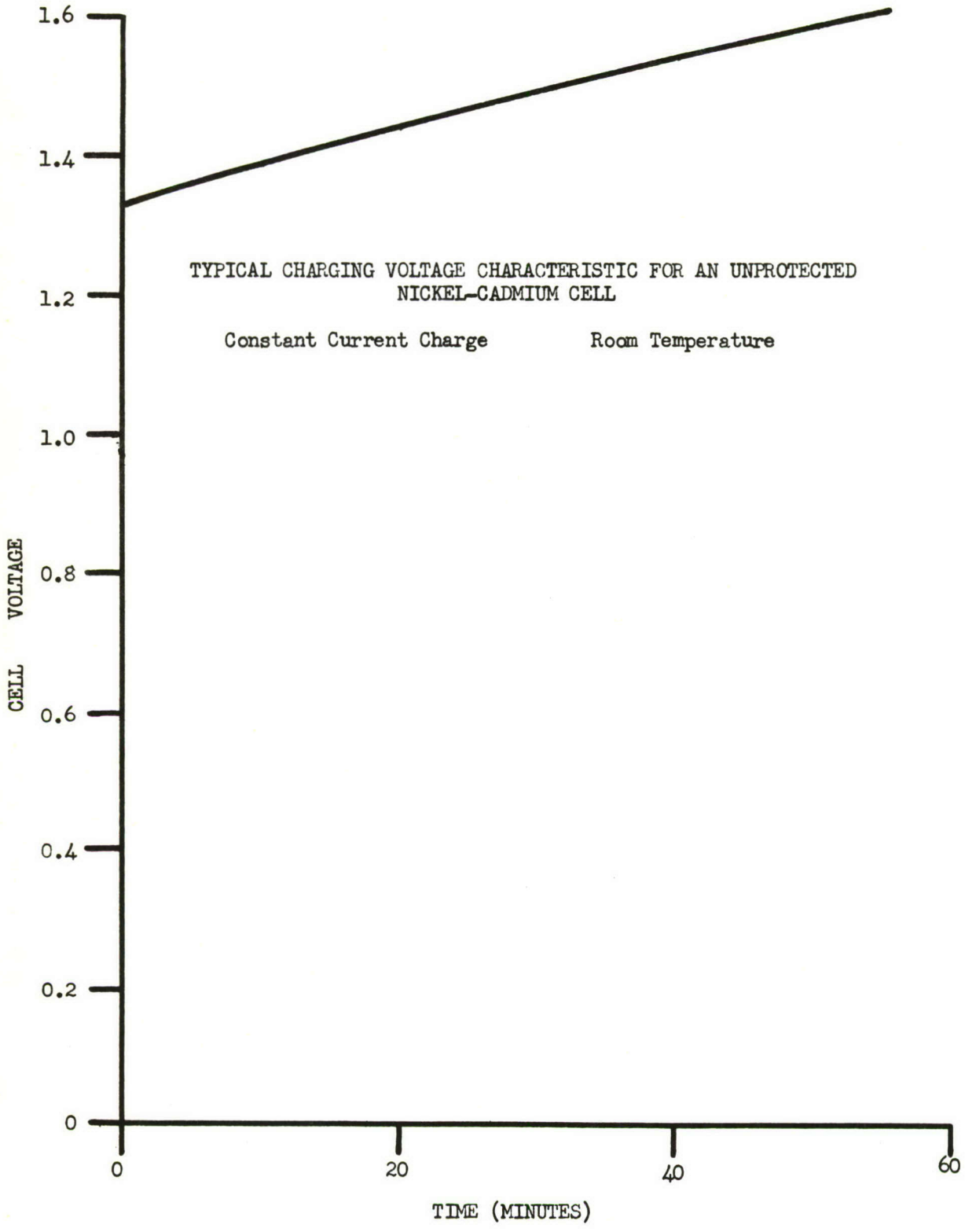


Figure 2



TYPICAL CHARGING VOLTAGE CHARACTERISTIC FOR AN UNPROTECTED NICKEL-CADMIUM CELL

Constant Current Charge Room Temperature

Figure 3

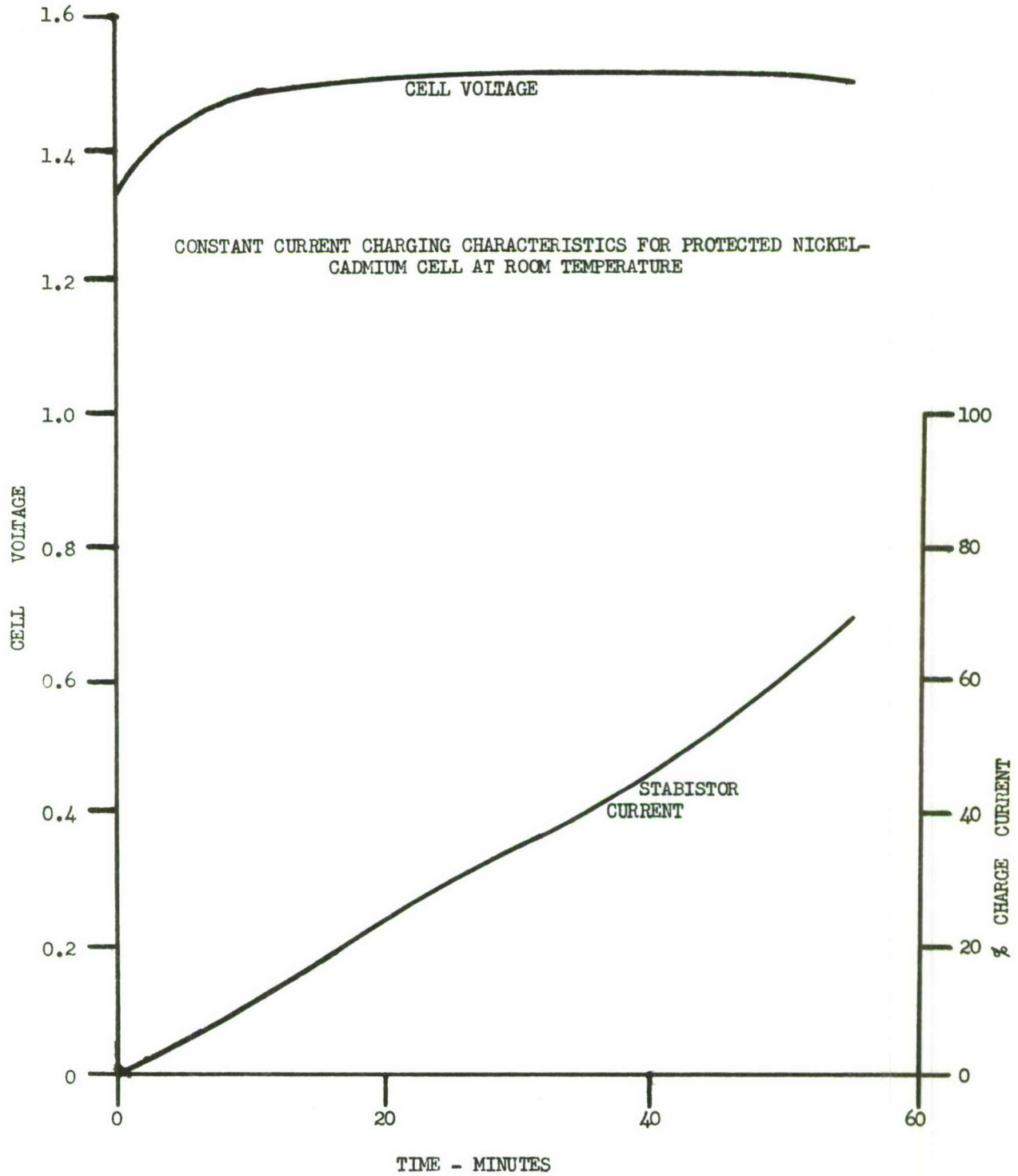
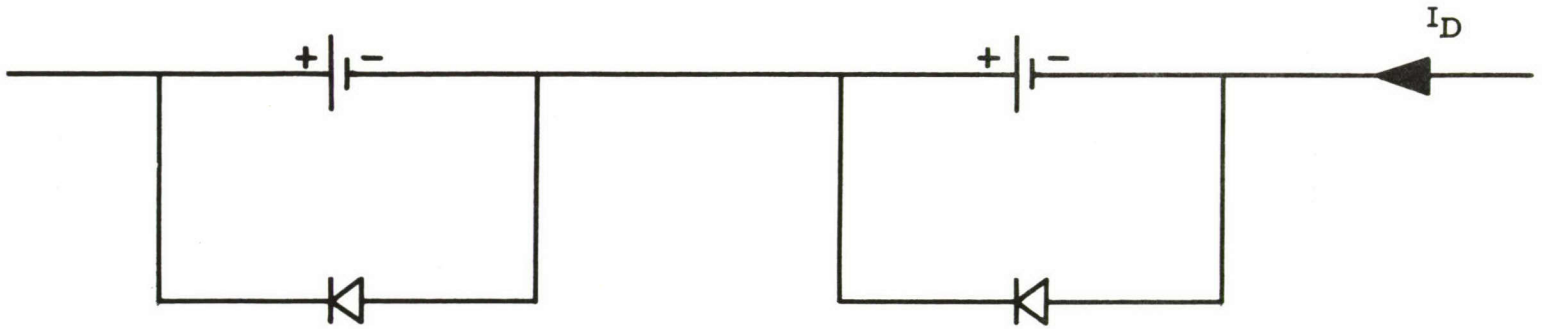

CONSTANT CURRENT CHARGING CHARACTERISTICS FOR PROTECTED NICKEL-CADMIUM CELL AT ROOM TEMPERATURE
CELL VOLTAGE
STABISTOR CURRENT
CELL VOLTAGE
% CHARGE CURRENT
TIME - MINUTES
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
100
80
60
40
20
0
0
20
40
60

Figure 4



"Diode for Control of Cell Polarity Reversal" (All polarities are shown for normal operation)

Figure 5

TYPICAL VOLT-AMPERE CHARACTERISTICS FOR IDEAL AND ACTUAL DIODES FOR CONTROL OF REVERSE CHARGING

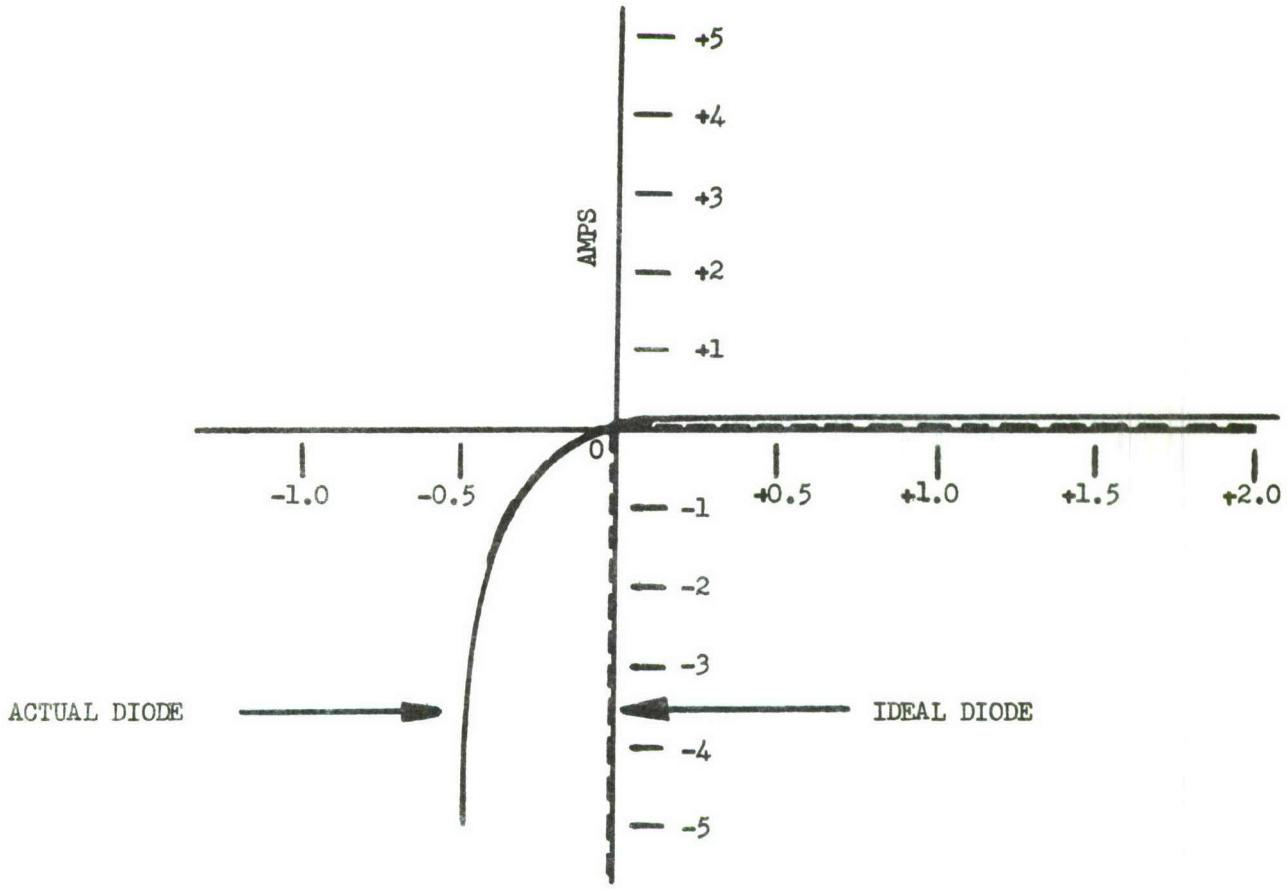
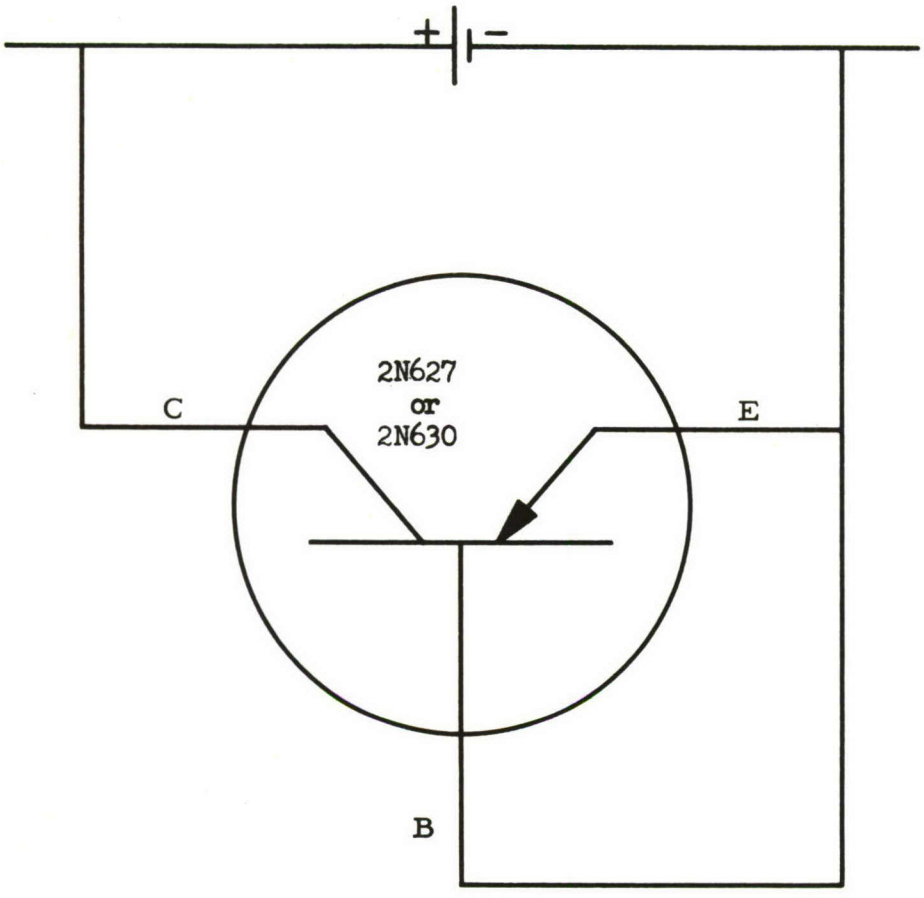
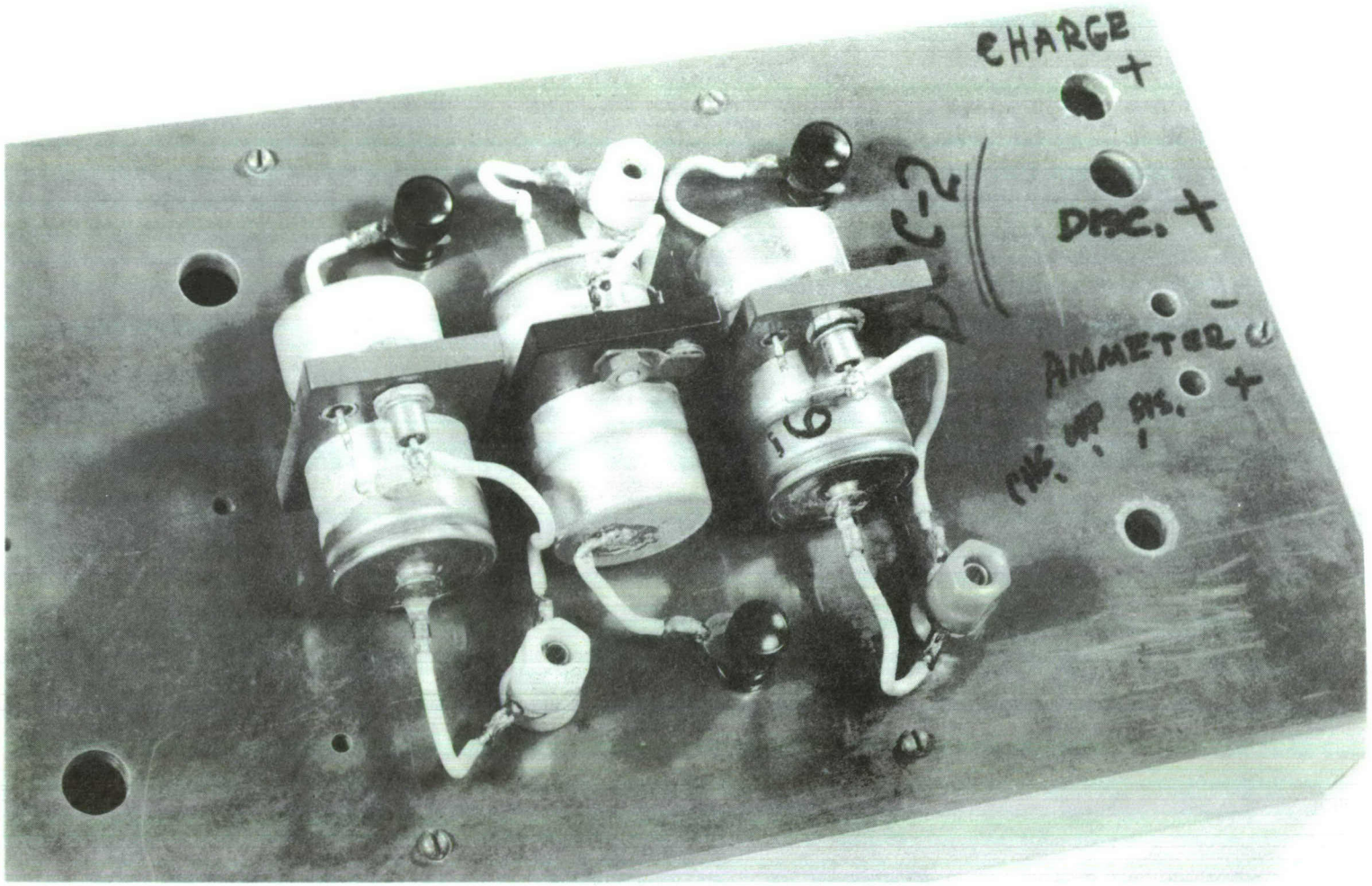


Figure 6



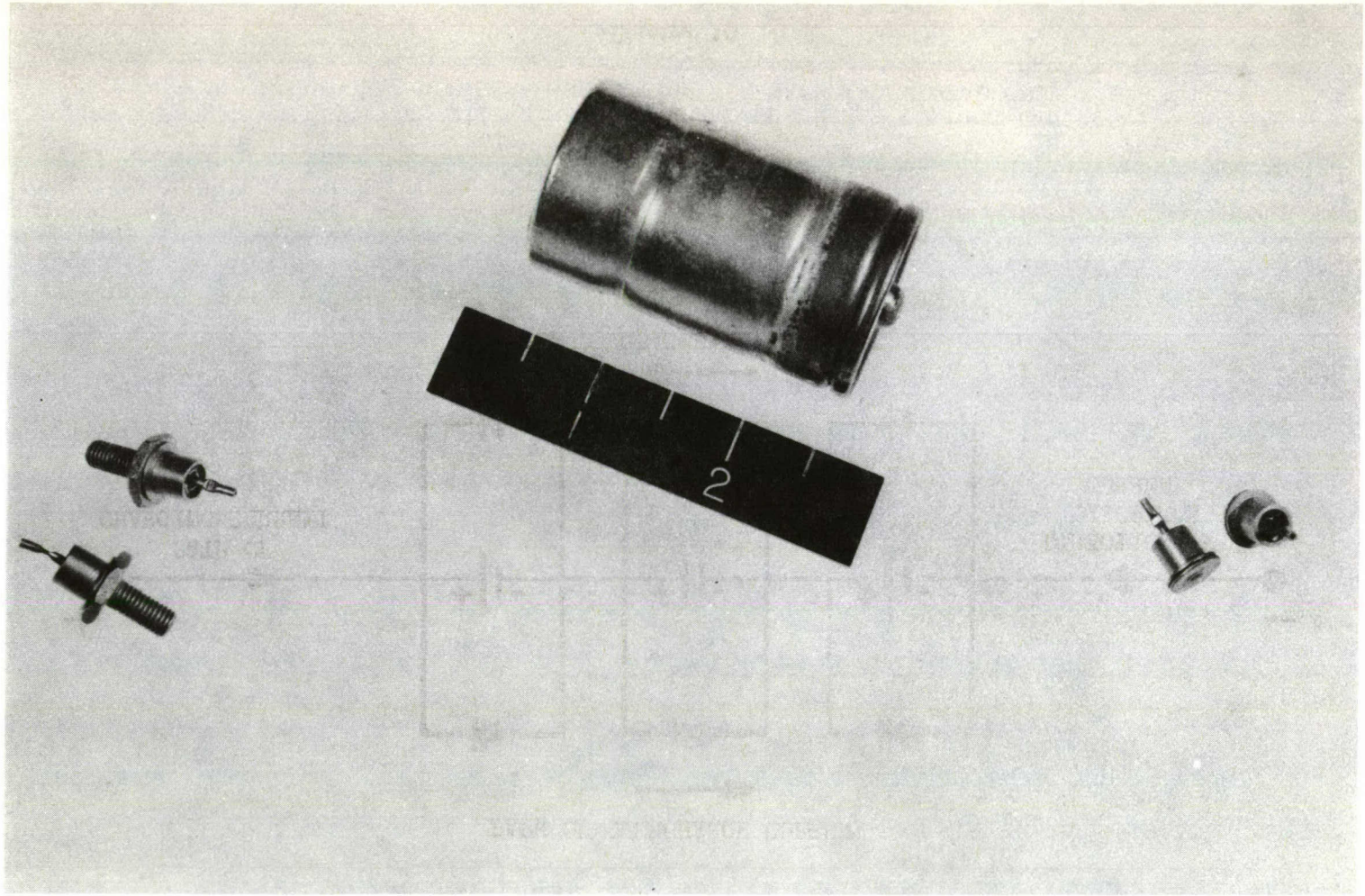
Power Transistor Connection for Control of Cell Reversal

Figure 7



THREE CELL BATTERY WITH ZENER DIODE END-OF-CHARGE VOLTAGE CONTROL AND DIODE CELL REVERSAL PROTECTION FOR EACH INDIVIDUAL CELL.

FIGURE 8



LOW VOLTAGE ZENER DIODE FOR END OF CHARGE CONTROL	3.4 AMPERE-HOUR "D" SIZE (FLASHLIGHT) NICKEL-CADMIUM BATTERY WITH COMMERCIAL TYPE CRIMP SEAL	HIGH FORWARD CONDUCTANCE DIODE FOR CELL REVERSAL PROTECTION DURING DISCHARGE

FIGURE 9

SCHEMATIC DIAGRAM FOR A THREE CELL BATTERY WITH DIODE CONTROL

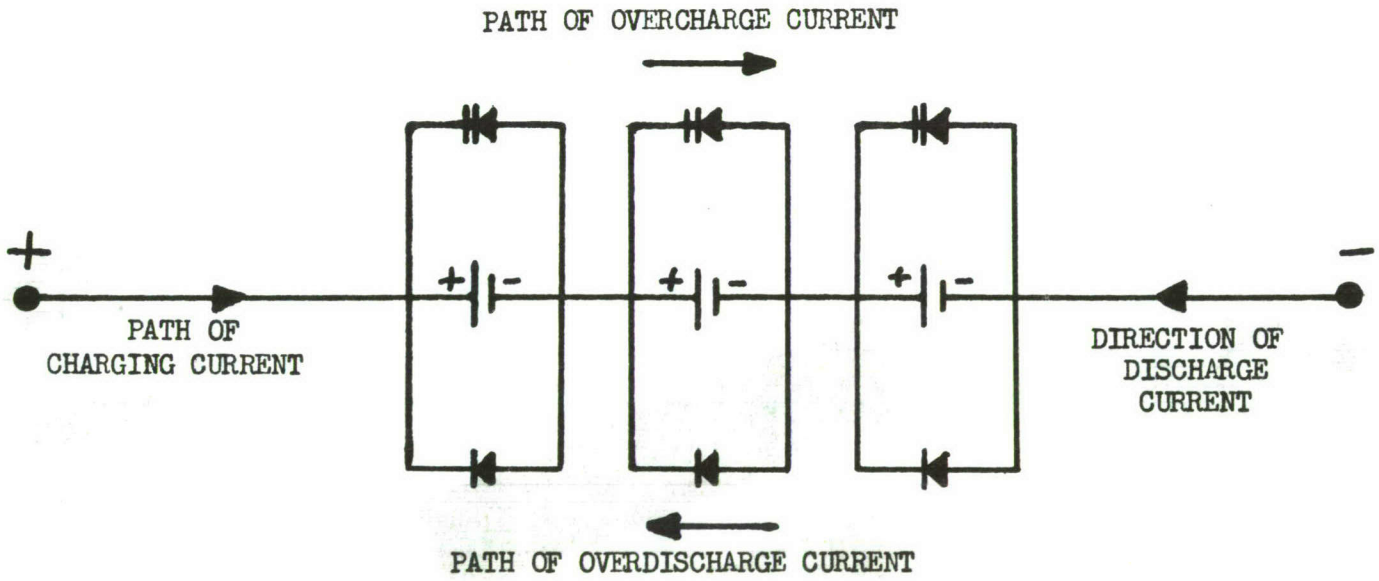


Figure 10

Terminal Voltage v.s. Charge-Discharge Cycles for 3.4 AH Nickel-Cadmium Cells

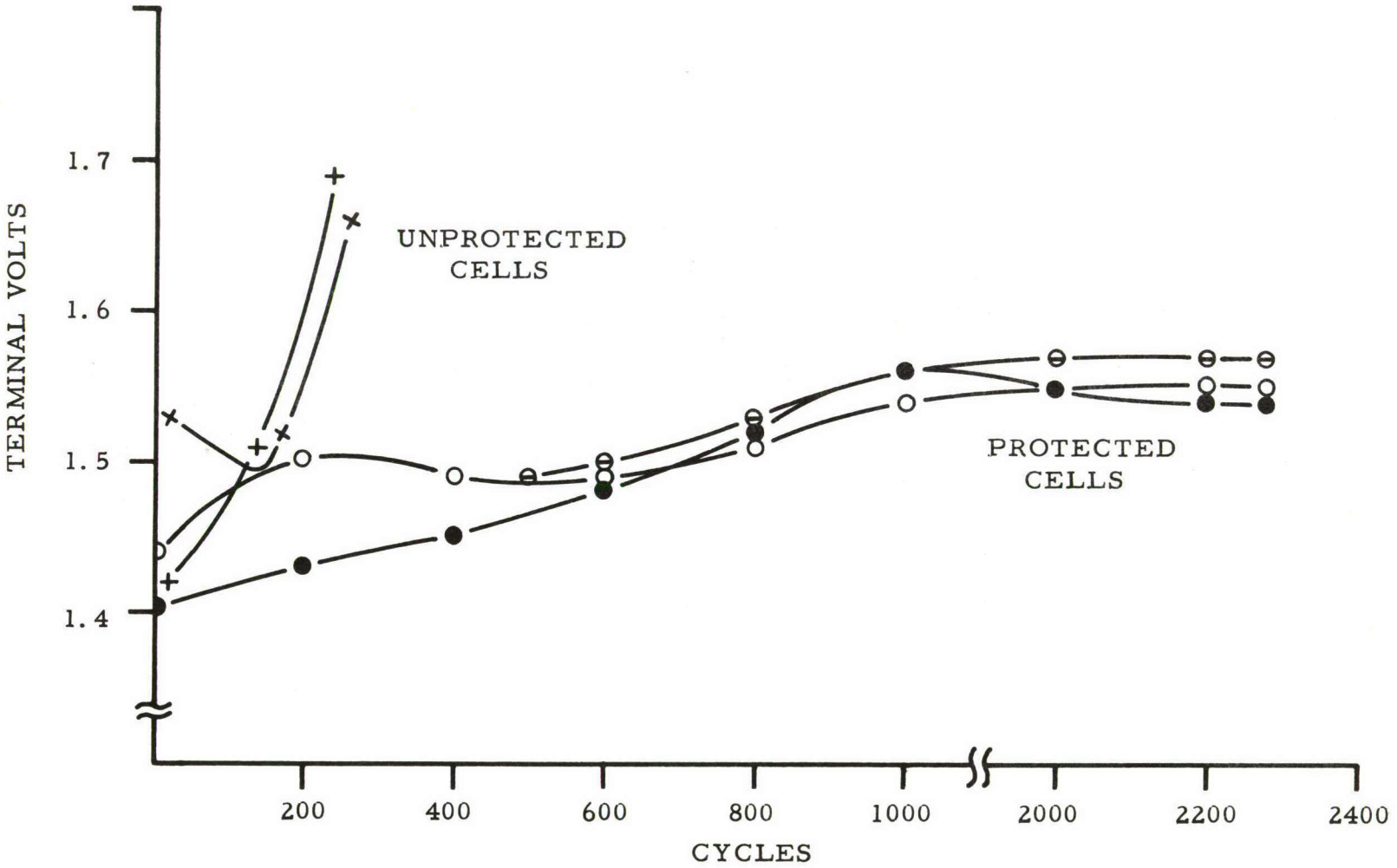


Figure 11

ACKNOWLEDGMENT

The author wishes to express his appreciation to Mr. James E. Cooper of the Air Force Aero Propulsion Laboratory for his assistance in the preparation of this paper and to P. R. Mallory and Company, especially Mr. F. J. Cocca, for the data furnished regarding diode construction, heat sinking of diodes and thermal coupling of the cell and diodes.

REFERENCES

1. Howard T. Francis, NASA Report SP-5004, "Space Batteries", Armour Research Foundation

2. W. W. Clark et al. RTD Technical Documentary Report 63-4144, "Alkaline Battery Evaluation", on Contracts AF33(616)-7529 and AF33(657)-8450, Inland Testing Laboratories

3. W. R. Cherry Paper, 17th Annual Power Sources Conference, "Some Characteristics of Spacecraft Solar Power Systems", National Aeronautics and Space Administration

4. Peter J. Whorisky, RTD Technical Documentary Report 63-4187, "Cell Equalization Techniques", on Contract AF33(657)-8749, P. R. Mallory and Company

5. W. S. Bishop, APIP Technical Memorandum 64-1, "Determination of a Three Cell Nickel Cadmium Storage Battery Performance Using Semiconductor Cell Equalization", Air Force Aero Propulsion Laboratory

6. F. J. Cocca, Paper, 18th Annual Power Sources Conference, "Stabistors - Their Application In Battery Operation", Contract Nr. AF33(615)-1342, P. R. Mallory and Company

HARMONIC GENERATION OF MICROWAVE PHONONS

Paul H. Carr

Air Force Cambridge Research Laboratories (OAR)

L. G. Hanscom Field, Bedford, Massachusetts

ABSTRACT

The ability to produce coherent phonons of microwave frequency provides a much needed means of investigating fundamental phenomena in solid state physics at frequencies never before attained. Microwave phonons were first produced in piezoelectric X-cut quartz in 1958. This paper shows for the first time that second and third harmonics of the fundamental acoustic wave are also generated. The fact that the second-harmonic acoustic power was observed to be proportional to the square of the fundamental and the third-harmonic acoustic power proportional to the cube of the fundamental establishes the nonlinearity of the generation mechanism. Second-harmonic generation of longitudinal phonons in nonpiezoelectric Z-cut quartz has also been observed. The generation mechanism is common to all dieletric media. This demonstrates that coherent phonon generation is not restricted to relatively rare piezoelectric materials as was commonly supposed. The highest frequency at which microwave phonons can be generated is limited to 80 Gcps by present-day high-power microwave sources. The successful generation of harmonics of microwave phonons enhances the possibility of extending the present limit well into the millimeter range. Fundamental research of this type can lead to such practical control devices as microwave delay lines since the acoustic delay time of one centimeter of quartz is the same as that of 500 meters of waveguide.

INTRODUCTION

The ability to generate coherent acoustic waves or phonons of microwave frequency provides a new and much needed means of investigating phenomena that are fundamental to solid state physics. A phonon is a quantum of the energy of vibration of the crystalline lattice analogous to a photon, which is an energy quantum of the electromagnetic field. Vibrations of the crystalline lattice are also called elastic or acoustic waves. Microwave phonons then, are elastic or acoustic waves having frequencies in the microwave, kilomegacycle, or gigacycle regions.

Microwave phonons up to frequencies of 2 Gcps were first generated in 1958 by Baranskii[1] and Bõmmel abd Dransfeld.[2] Jacobsen[3] has generated phonons to as high as 24 Gcps, and Thaxter and Tannenwald[4] have recently gone as high as 70 Gcps. The phonons were generated by placing a piezoelectric quartz rod in the high-intensity electric field region of a reentrant cavity. The present paper shows that second and third harmonics of the fundamental acoustic wave are also generated at the same time; the second harmonic is large enough to be of significance in harmonic-generating phonons in the millimeter range, which is currently inaccessible due to the lack of high-power microwave sources. Second-harmonic generation of longitudinal phonons in nonpiezoelectric Z-cut quartz was for the first time observed in this experiment. The generation mechanism is apparently common to all dielectric media. This indicates that coherent acoustic wave generation is not restricted to the relatively rare piezoelectric materials as has commonly been supposed. The present experiments were performed between 1.5° to 4.2° K to eliminate the high attenuation caused

by the thermal scattering of microwave phonons at higher temperatures.

Fundamental research of this type is giving rise to a new technology for the control of microwave radiation. The devices are characterized by their compactness, a necessary feature for space applications. Since the velocity of sound is many orders of magnitude smaller than the velocity of light, an acoustic delay line 1 cm in length can give the same time delay as 50,000 cm of waveguide. Delay lines made of magnetic materials have been made so that the delay time can be varied by changing the magnetic field.[5] Gains of 30 db/cm do not appear to be unreasonable for traveling-wave acoustic amplifiers in piezoelectric semiconductors. Phonon-pumped parametric amplifiers and nonreciprocal acoustic devices will also be available.

THE EXPERIMENT

Sound waves are generated at the end surface of a piezoelectric quartz rod placed in the high-intensity electric field region of a reentrant 4.5-Gcps cavity shown in Figure 1. This cavity is excited by a 0.5-μsec pulse of electromagnetic energy passed through a low-pass filter to eliminate any harmonics above 4.5 Gcps that may be present in the pulsed microwave source. The longitudinal hypersonic sound waves or microwave phonons propagate with a velocity of 5.7×10^5 cm/sec to the other end of the rod, which is placed in a 9-Gcps cavity. There the second-harmonic content of the acoustic wave is partially converted into microwave energy by the inverse piezoelectric effect and the resulting signal from this cavity is detected by a superheterodyne receiver having a minimum sensitivity of about 10^{-12} watts. The acoustic wave reflected from this end returns to the generating cavity where the fundamental acoustic energy is detected in a

similar manner by a second receiver. The square root of the ratio of the microwave signal from the first acoustic echo to the microwave power absorbed by the 4.5-Gcps cavity measures the electromechanical conversion efficiency of the rod in this cavity,[6] which is on the order of 10^{-4}. The series of echoes generated and detected in the 4.5-Gcps cavity as the wave packet undergoes successive reflections at the end faces of the rod is shown in Figure 2(a), and the second harmonic detected in the 9-Gcps cavity is shown in Figure 2(b). The beat structure superposed on the exponential decay is caused by interference effects due to the nonparallelism of the end faces.[3, 6] The end faces here are parallel to 1.5 sec of arc. (The wavelength of sound at 4.5 Gcps is 1.3 μ.) The interference is consistent with the fact that the wavelength of the harmonic is one-half that of the fundamental.

In the second part of the experiment, microwave acoustic waves are generated by a pulse of electromagnetic radiation in the 9-Gcps cavity, and the resulting echoes, shown in Figure 2(c), are detected in the same cavity. The electromechanical conversion efficiency for the rod in this cavity, which is also on the order of 10^{-4}, is measured in the same manner as in the 4.5-Gcps cavity.

A plot of the microwave power of each echo as a function of distance traveled is shown in Figure 3. This is measured by introducing a calibrating pulse of known power into the receiver at a slightly earlier or later time than the arrival of an echo. It can be seen that the phonons generated at the end surface of the rod in the 9-Gcps cavity initially fall off at the same rate as the harmonic content of the wave generated in the 4.5-Gcps cavity. From this it is concluded that the nonlinearity causing

the harmonic generation occurs at the end surface of the quartz rod. If the nonlinearity had been due to nonlinearity of the elastic medium, then the harmonic content of the wave would have increased, come to a maximum at a distance about that of the phonon mean free path, and decayed exponentially.[5] Since the mean free path is about 10 cm and the length of the rod 1 cm, any nonlinearity occurring in the volume of the rod would have been observed.

The third harmonic is generated by replacing the 4.5-Gcps cavity, which has a gap spacing of 0.79 mm, by a 3-Gcps reentrant cavity, which has a gap of 1.65 mm and other dimensions chosen to resonate at the desired frequency. In Figure 4, the odd echoes that are aligned with the vertical graticules are the third-harmonic content of the wave generated in the 3-Gcps cavity and detected in the 9-Gcps cavity. The even-numbered echoes are generated in the 9-Gcps cavity at the same time that the 3-Gcps cavity is excited and their amplitude is adjusted to give the even-decay pattern seen. It can be seen that except for the number four third-harmonic echo, which is anomalously large owing probably to interference effects, the third harmonic decreases at the same rate as 9-Gcps phonons generated in the 9-Gcps cavity. A measurement of the amplitude of each echo produced a curve similar to that of Figure 3, and it may again be concluded that the nonlinearity occurs at the surface of the quartz rod.

THEORETICAL CONSIDERATIONS

The physical process causing the harmonic generation in piezoelectric quartz could be accounted for in terms of the higher-order piezoelectric constants defined by Mason.[7] Concerning the nonlinear behavior of the piezoelectric constant, the only experimental work known to the author is that of Tsi-Ze,[8] who observed that when opposite sides of an X-cut slab

of quartz are covered with metal electrodes and a large static potential difference V placed across them, the resulting strain can be expressed by

$$S = \frac{\Delta e}{e} = \pm \frac{d_{11}}{c_p e} [\, 1 - \exp(-c_p V) \,] - c_d \frac{V^2}{e^2} , \qquad (1)$$

where $d_{11} = 6.4 \times 10^{-8}$ is the low-field piezoelectric constant, e is the thickness of the plate, $c_p = 1/114$, and $c_d = 5.5 \times 10^{-12}$ in esu. The first term expresses the field dependence of the piezoelectric constant; the second, the strain on any dielectric when it is placed in an electric field. This term is much too small to account for the observed value of the second harmonic in piezoelectric quartz, and will therefore be disregarded. This is in agreement with some preliminary observations of Quate[9] but in disagreement with those of Shiren.[10] The plus sign in the first term refers to a positive voltage across the slab, and the negative sign refers to a negative voltage. The voltage V in the exponential is to be regarded as an absolute value. This term therefore has odd symmetry, and could at best be related only to the third harmonic.

Let us ask whether, independently of the work of Tsi-Ze, the results reported in this paper can be fitted to an exponential field dependence of the piezoelectric constant, that is, to $d(E) = d_{11} \exp(-kE)$, where k is the saturation constant to be determined from a fit to the data. Expanding the exponential yields

$$S = d_{11}E - \frac{d_{11}k}{2} E^2 + \frac{d_{11}k^2}{6} E^3 . \qquad (2)$$

It can be shown[3, 6] that the wave equation in a piezoelectric medium is

$$\frac{\partial^2 u}{\partial x^2} - \frac{1}{v_{11}^2} \frac{\partial^2 u}{\partial t^2} = \frac{\partial}{\partial x} S = S\,\delta(x - x_o) . \qquad (3)$$

Here u is the displacement of a macroscopic volume element, v_{11} is the velocity of a longitudinal wave along the X axis, x_o is the position at the end of the rod, and $\delta(x - x_o)$ is the delta function. The delta function comes in because the strain undergoes a discontinuity at the end surface.

If a sinusoidally varying electric field is present at the end of the quartz, the first term in Eq. 2 generates the fundamental; the second, the second harmonic; and the third, the third harmonic. With the appropriate Green's function, Eq. 3 can be solved for u, which in turn is used to compute the acoustic power P_f, P_{2f}, P_{3f} at the fundamental and the second and third harmonics respectively. Thus,

$$P_f = \left(\frac{1}{2} c_{11} v_{11} d_{11}^2 A\right) E_r^2 ; \tag{4a}$$

$$P_{2f} = \frac{1}{16} \left(\frac{1}{2} c_{11} v_{11} d_{11}^2 A\right) k^2 E_r^4 ; \tag{4b}$$

$$P_{3f} = \frac{1}{576} \left(\frac{1}{2} c_{11} v_{11} d_{11}^2 A\right) k^4 E_r^6 . \tag{4c}$$

Here c_{11} is the elastic constant, E_r is the electric field just inside the end of the quartz rod and is assumed to be uniform; A is the area of the rod.

The following ratios are useful for experimental comparison:

$$\frac{(P_{3f}/P_f^3)^{\frac{1}{2}}}{P_{2f}/P_f^2} = \frac{2}{3} ; \tag{5a}$$

$$\frac{P_{2f}}{P_f^2} = \frac{k^2}{16(\frac{1}{2} c_{11} v_{11} d_{11}^2 A)} ; \tag{5b}$$

$$\frac{P_{3f}}{P_f^3} = \frac{k^4}{576(\frac{1}{2} c_{11} v_{11} d_{11}^2 A)^2} \tag{5c}$$

Is it possible to relate the empirical constant k to the observations of Tsi-Ze? If the relation V = Ee valid for the parallel-plate capacitor configuration of Tsi-Ze is used in Eq. 1, $k = c_p e$. If the equivalent circuit approximation for a reentrant cavity is used, the high-intensity electric field gap where the rod is placed can be replaced by a lumped capacitance of the same dimensions as the gap. The thickness of the gap is equivalent to the thickness of the plate used by Tsi-Ze, and where this is about 1 mm in the experimental cavities, the value of c_p measured by Tsi-Ze yields $k = 2 \times 10^{-8} mv^{-1}$. At the end surface of the rod the acoustic waves are generated to a depth of about an acoustic wavelength, which is 10^{-3} mm. This would make the predicted value of k three orders of magnitude smaller.

The ambiguity as to which thickness to use in $k = c_p e$ arises because the empirical relation in Eq. 1 is a function of the voltage V and not of the electric field E. This empirical relationship, which has been criticized by Cady[11] as not being on firm physical grounds, therefore prevents a quantitative comparison between the results of the present experiment and those of Tsi-Ze.

DISCUSSION

The technique for selecting the rods represented by the data in Table 1 was as follows. The parallelism was determined to be better than 6 sec of arc, first by inspection at room temperature with an optical autocollimator, and then by examination of the beat pattern at 4.2° K. By measuring the electric field by a perturbation technique,[12] the combination of elastic constants and piezoelectric constants in Eq. 4a was found to be equal to the values obtained by Mason[7] to within an estimated 50% experimental

error, the main source of error being the spread in the measured values of the electromechanical efficiency. Two of three X-cut rods and one of two AC-cut rods, all of about 3 mm in diameter and 1 cm in length, met these criteria.

The acoustic power P was computed from a measurement of the microwave power and the electromechanical conversion efficiency. The electric breakdown across the gap of the reentrant cavity placed an upper limit on the acoustic power. For the 4.5-Gcps cavity this was about 20 mw; for the 3-Gcps, it was about 100 mw. For the maximum acoustic power of X-cut quartz at 4.5-Gcps, the ratio P_{2f}/P_f was 4×10^{-2}; at 3-Gcps, $P_{3f}/P_f = 2 \times 10^{-7}$.

The second-harmonic power P_{2f} was found to be proportional to the square of the fundamental power; the third P_{3f} was proportional to the cube of the fundamental. The measured values of P_{2f}/P_f^2 and P_{3f}/P_f^3 are tabulated in Table 1. The spread reflects the reproducibility of the data. The value of P_{2f}/P_f^2 was found to be reproducible to within ± 90% over six different runs with the same X-cut rod. This spread is diminished by the square root in the computation of k. One run consisted in mounting the rod in the cavities, cooling the assembly to 1.4° K, taking the measurements, and warming up the assembly to room temperature. There was no change in P_{2f}/P_f^2 within the experimental reproducibility when the ends of the rods between the two cavities were reversed. No temperature variation in P_{2f}/P_f^2 was observed in the range from 1.4° to 4.2° K.

The values of k in Table 1 were computed for X-cut quartz from Eq. 5b for the second harmonic and from 5c for the third, using the measured value

of the acoustic power and the value given by Mason[7] for the elastic and piezoelectric constants. A similar procedure was followed for the AC-cut rods.

The experimental value of the ratio

$$\frac{(P_{3f}/P_f^3)^{\frac{1}{2}}}{P_{2f}/P_f^2}$$

for transverse waves in AC-cut quartz is equal within experimental error to the value predicted for an exponential saturation of the piezoelectric constant. The values of k computed from the second and third harmonics are equal within the reproducibility of the data, as they should be. In addition they agree in order of magnitude with the value of $k = 2 \times 10^{-8} mv^{-1}$ from the previously mentioned experiments of Tsi-Ze. This agreement could be fortuitous since the validity of using the gap size of the cavity, which is necessary for a comparison with the experiments of Tsi-Ze, has not been proved experimentally.

The experimental value of the ratio

$$\frac{(P_{3f}/P_f^3)^{\frac{1}{2}}}{P_{2f}/P_f^2}$$

clearly does not fit the theory for X-cut quartz; however, the value of k computed from the third harmonic is of the same order of magnitude as that of AC-cut quartz. It is therefore possible that the generation mechanism for the third harmonic in X-cut quartz is the exponential saturation of the piezoelectric constant. The value of k for the second harmonic in X-cut quartz is an order of magnitude too large. Some other generation mechanism may be present here.

Second-harmonic generation of longitudinal waves in nonpiezoelectric Z-cut quartz has also been observed. The fact that no echoes were received in the 4.5-Gcps cavity demonstrates that the generation mechanism has no inverse. The harmonic echoes shown in Figure 5 were detected at 9 Gcps by evaporating a thin piezoelectric film of cadmium sulphide on the end of the rod in the 9-Gcps cavity. Quantitative results show that the second-harmonic power is about five orders of magnitude smaller than the fundamental in X-cut piezoelectric quartz and that the generation mechanism is due to the discontinuity in the Maxwell stress tensor. This result confirms that the generation process is not White's[13] dielectric-heating mechanism, which would give the same order of magnitude for both Z-cut and X-cut quartz. It also indicates the possibility of harmonically generating acoustic waves in any nonpiezoelectric media.

CONCLUSIONS

Second- and third-harmonic generation of transverse waves in AC-cut quartz and longitudinal waves in X-cut quartz has been observed. The relative magnitudes of the second and third harmonics in AC-cut quartz can be fitted to an exponential saturation of the piezoelectric constant. This is not true for longitudinal waves in X-cut quartz, the second harmonic being too large.

The second harmonic in X-cut quartz is large enough to be of use in extending present techniques to the millimeter microwave range. The highest frequency at which high-power pulsed microwave sources are available is 80 Gcps. Thus, the harmonic generation of 160-Gcps phonons is now shown to be possible. The maximum value of the fundamental acoustic power in this experiment was limited to 0.1 w by electric

breakdown in the cavity. Increasing the breakdown field could make the harmonic of the same magnitude as the fundamental for acoustic power of the order of 0.5 w.

Second-harmonic generation of longitudinal phonons in nonpiezoelectric Z-cut quartz has also been observed, the generation mechanism being apparently common to all dielectric media. Thus, contrary to what has commonly been supposed, coherent phonon generation may not be restricted to the relatively rare piezoelectric materials.

ACKNOWLEDGMENTS

I am indebted to Professor S. Berko of Brandeis University for helpful discussions, Dr. J. deKlerk for fabricating the thin film CdS transducer, W.M. Lepsevich for constructing the cavities, and J. Silva for technical assistance.

TABLE 1. Experimental Data

Polarization	2d Harmonic		3d Harmonic		$\frac{(P_{3f}/P_f^3)^{\frac{1}{2}}}{P_{2f}/P_f^2}$	
	$P_{2f}/P_f^2 w^{-1}$	$k\left(\frac{m}{v}\right)$	$P_{3f}/P_f^3 w^{-2}$	$k\left(\frac{m}{v}\right)$	Experiment	Theory
Longitudinal (X-cut)	0.84-2.7	$3.2\text{-}5.8 \times 10^{-7}$	$0.82\text{-}1.9 \times 10^{-5}$	$2.4\text{-}3 \times 10^{-8}$	$1.6\text{-}3.3 \times 10^{-3}$	0.667
Transverse (AC-cut)	$1.9\text{-}3.2 \times 10^{-2}$	$2.8\text{-}4.6 \times 10^{-8}$	2.4×10^{-4}	4.0×10^{-8}	0.48-0.79	0.667

REFERENCES

1. Baranskii, K.N., Soviet Physics, Doklady 2:237 (1958)
2. Bömmel, H.E., and Dransfeld, K., Phys. Rev. Letters 1:234 (1958)
3. Jacobsen, E.H., Phys. Rev. Letters 2:249 (1959); J. Acoust. Soc. Am. 32:949 (1960); Quantum Electronics (Columbia University Press, 1960)
4. Thaxter, J.B., and Tannenwald, P.E., PTGMTT Symposium (1964)
5. Olson, F.A., and Yaeger, J.R., Appl. Phys. Letters 5:33 (1964)
6. Carr, P.H., and Strandberg, M.W.P., J. Phys. Chem. Solids 23:923 (1962)
7. Mason, W.P., Physical Acoustics and the Properties of Solids (Van Nostrand, 1958)
8. Tsi-Ze, Ny, Compt. rend. 184:1645 (1927); J. Phys. Radium 9:13 (1928)
9. Quate, C.F., Stanford Q. Status Rpt. No. 20, Contract No. 225(48) Code No. 373-361, April 1964
10. Shiren, N.S., IBM Quarterly Progress Report, Phonon Interactions in Crystals, Contract No. DA36-039AMC-02280(E), Aug. 1963
11. Cady, W.G., Piezoelectricity, McGraw-Hill, 1946, p. 221
12. Carr, P.H. (to be published)
13. White, R.M., J. Appl. Phys. 34:3559 (1963)

Figure Legends

1. Cross-sectional view of experimental setup.
2. Oscilloscope traces of longitudinal echoes in X-cut quartz: (a) 4.5-Gcps fundamental, scale 20 μsec/main scale division; (b) second harmonic at 9-Gcps, scale 10 μsec/div; (c) echoes generated and detected in 9-Gcps cavity, scale 20 μsec/cm. The first pulse on the left in (a) and (c) is the generating electromagnetic pulse. In (b), pickup from the generating pulse is to the left of the edge of the photograph.
3. The microwave power of each echo shown in Figure 2, measured as a function of distance traveled.
4. Third-harmonic transverse echoes in AC quartz are aligned with the main scale division. Even-numbered echoes were generated in the 9-Gcps cavity and their amplitude adjusted to give the even decay mode seen.
5. Second-harmonic generation in nonpiezoelectric Z-cut quartz. The harmonics were generated in the 4.5-Gcps cavity by the Maxwell-Faraday stress and detected in the 9-Gcps cavity by means of a thin-film CdS piezoelectric transducer evaporated onto this end of the rod. The scale is 2 μsec/cm. The time between the first pulse, which is caused by pickup from the microwave source, and the first echo corresponds to the time that it takes for the acoustic wavepacket to travel one length of the rod.

Figure 1.

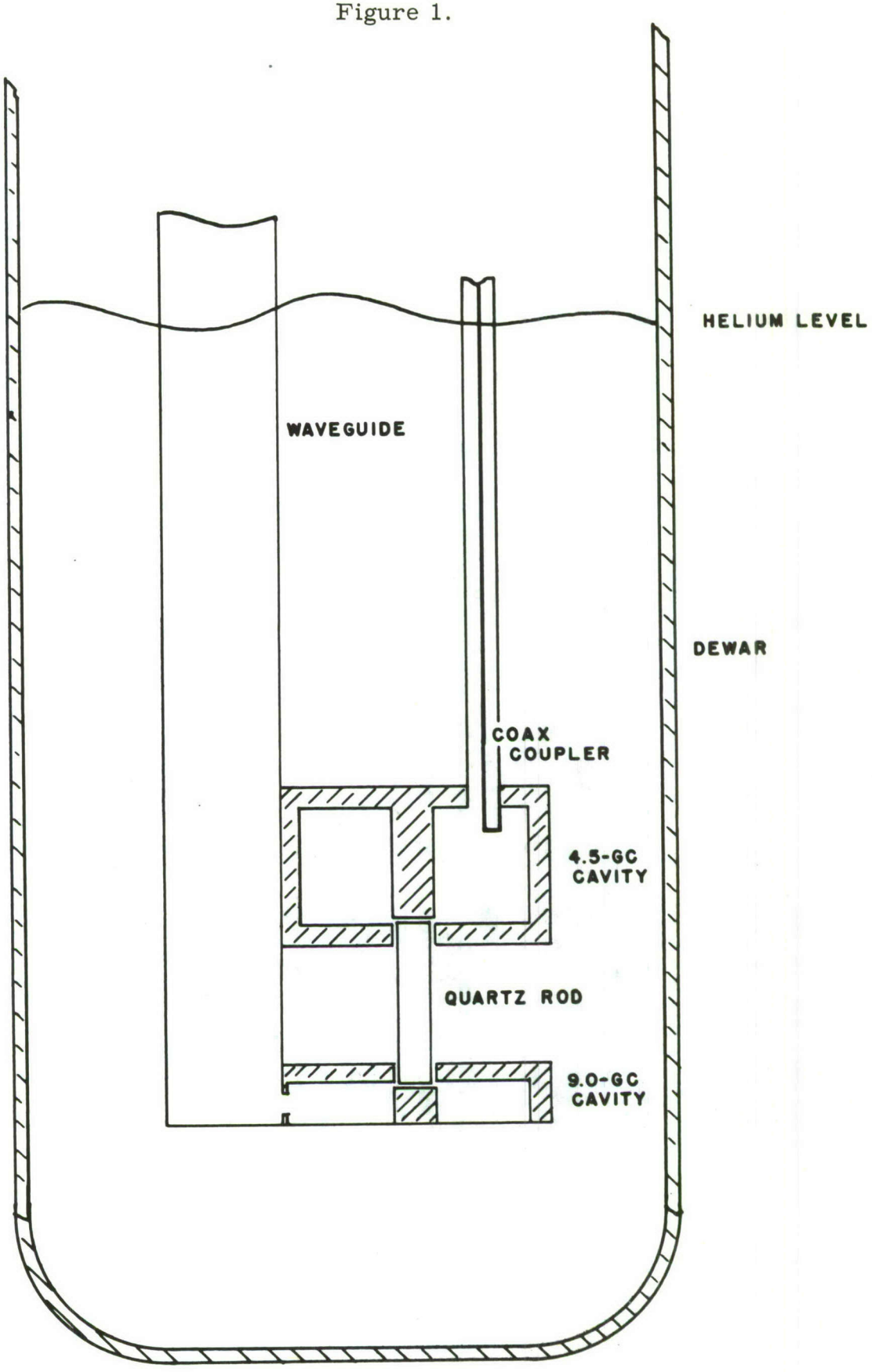
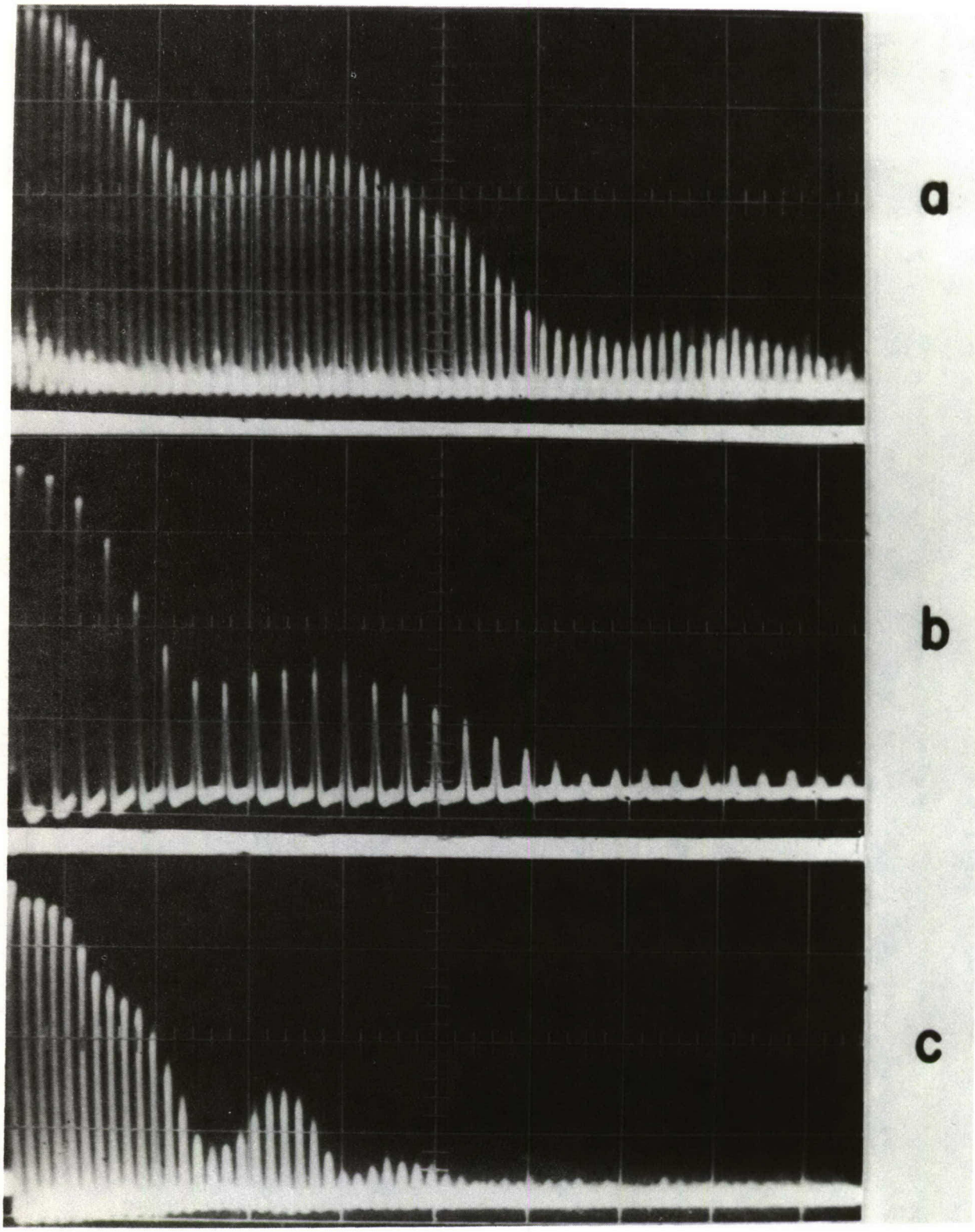

HELIUM LEVEL
WAVEGUIDE
DEWAR
COAX
COUPLER
4.5-GC
CAVITY
QUARTZ ROD
9.0-GC
CAVITY

Figure 2.



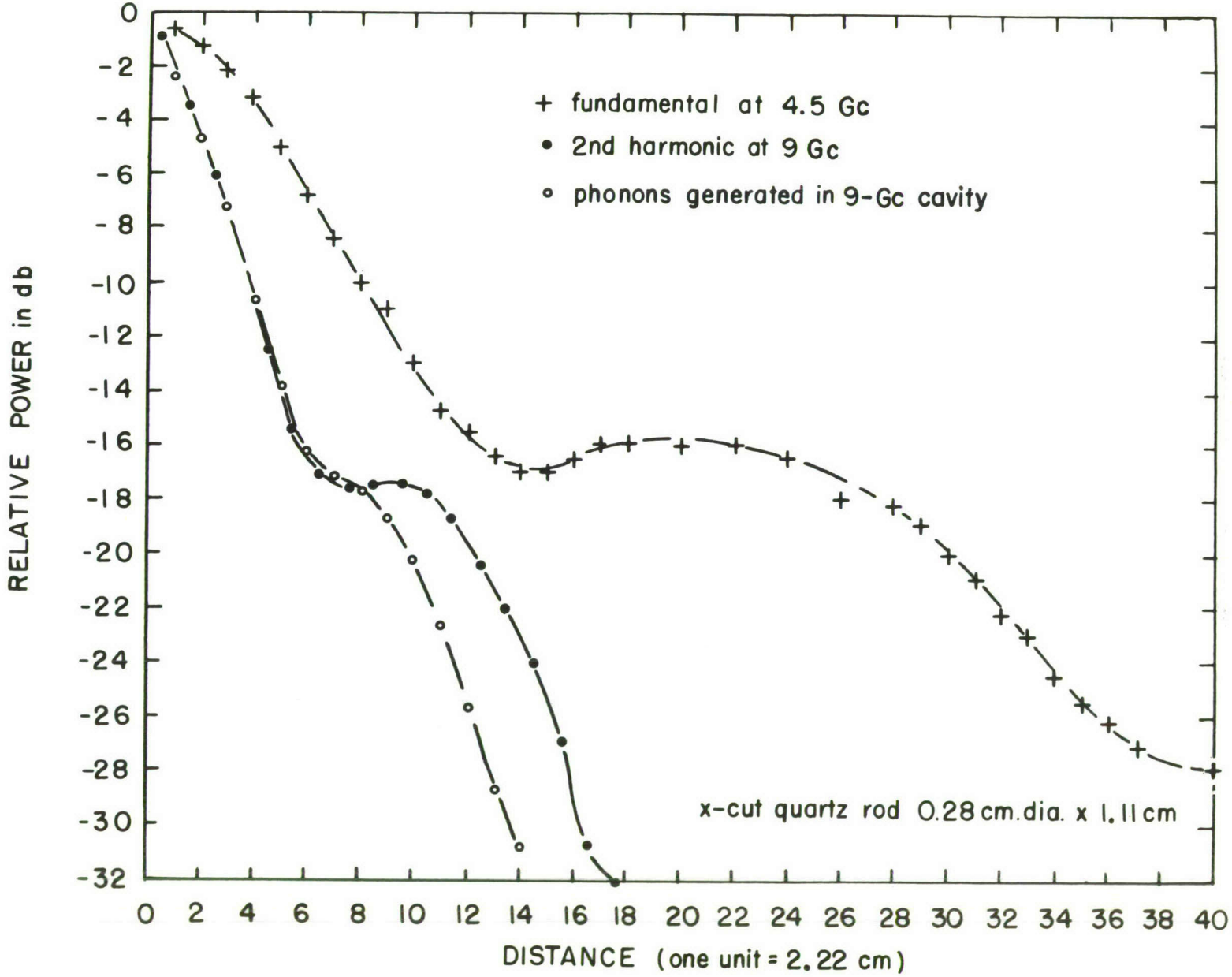

RELATIVE POWER in db
0
-2
-4
-6
-8
-10
-12
-14
-16
-18
-20
-22
-24
-26
-28
-30
-32
+ fundamental at 4.5 Gc
• 2nd harmonic at 9 Gc
° phonons generated in 9-Gc cavity
x-cut quartz rod 0.28 cm. dia. x 1.11 cm
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40
DISTANCE (one unit = 2.22 cm)

Figure 3.

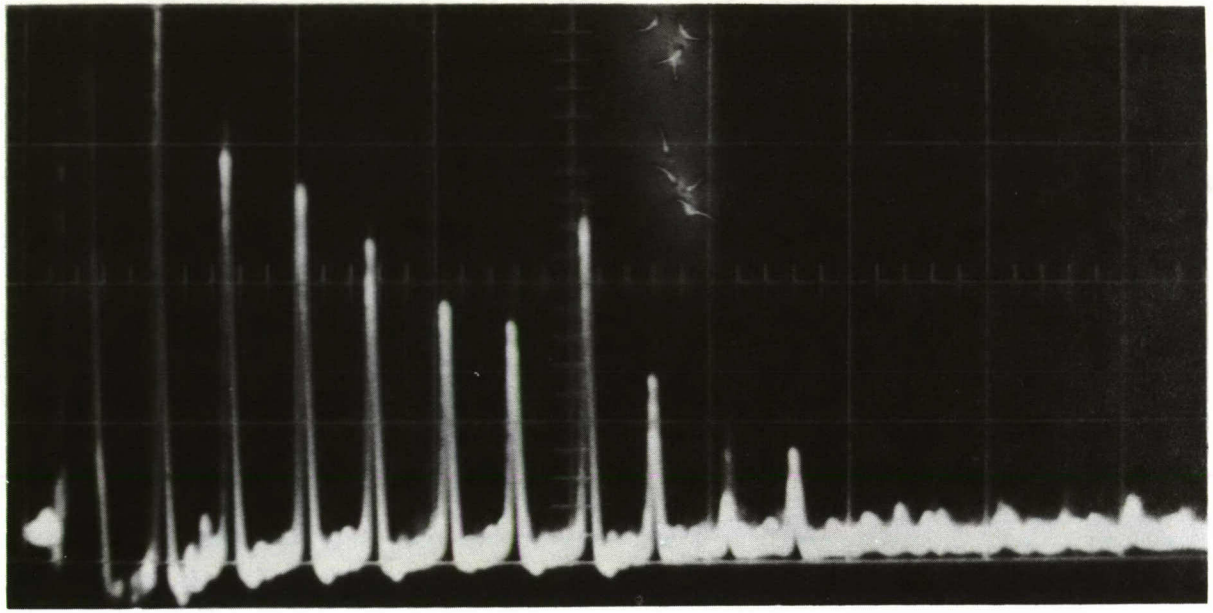


Figure 4.

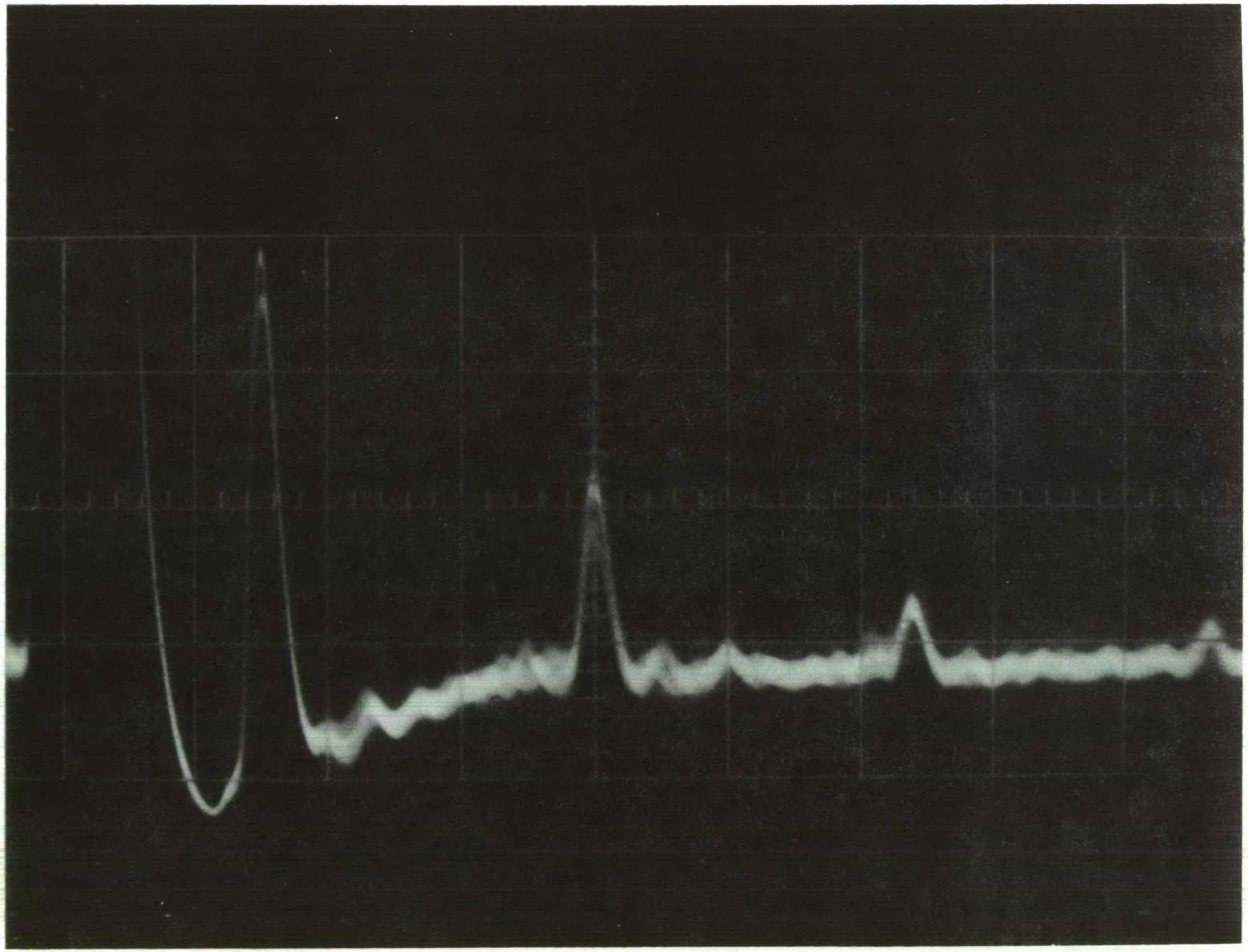


Figure 5.

SPECTRA OF ANIONS: ACETYLACETONATES

Carl A. Pitha
Air Force Cambridge Research Laboratories (OAR)
L. G. Hanscom Field, Bedford, Massachusetts

Although the spectra of transition and rare earth metal ions have been the object of much study, the spectra of anions have received only cursory consideration. Since the electronic states of anions may participate importantly in the optical absorption of materials useful in quantum electronic devices, an investigation was undertaken to establish whether the electronic states of an anion are influenced by an associated metal ion, and if they are, under what conditions.

An investigation of the ultraviolet absorption spectra of more than 40 metal acetylacetonates in iso-octane solution reveals that only certain metal ions influence the electronic states of metal acetylacetonates, and that the critical characteristic of the metal ion is its polarizing ability as reflected in Z/r. Water as a solvent, contaminant, or component of the acetylacetonate, as in hydrates, destroys this interaction. Other metal ion characteristics which appear as significant variables are the presence or absence of d-orbital electrons, valence, magnetic state, and the molecular geometry. The perturbing influence of "reactive" solvents, such as the alcohols, is sufficiently great to obliterate most, if not all relationships.

SPECTRA OF ANIONS: ACETYLACETONATES

INTRODUCTION

Development of new quantum electronic devices, such as LASER, for Air Force and commercial technological applications emphasizes the need to know what factors control and influence the optical absorption (and therefore, the electronic states) of ions in single crystals, glasses, solutions, and gases which are the active elements in these devices. The crystal and ligand field theories have recently provided guidance and insight in increasing this understanding, particularly of the influence of the symmetry and intensity of the field surrounding a transition metal ion on its spectrum. The spectra of anions (negatively charged ions, such as Cl^-, NO_3^-, $SO_4^=$, $PO_4^{\equiv}$) on the other hand, have received only cursory consideration. There are few extensive investigations, for instance, of the influence of association with a wide range of metal ions on the spectrum of an anion, and most of the conclusions of these investigations must be seriously questioned. To provide a firmer basis for evaluating the electronic spectra of materials potentially useful in quantum electronic devices, a study was undertaken as part of the program of the Solid State Sciences Laboratory of AFCRL to establish:

1. Whether the electronic states of the anion are influenced by the associated metal ion.

2. Under what conditions, if any, the electronic states of the anion are affected by such association.

3. Whether any patterns of interdependence of spectral, physical and chemical properties can be established.

A major problem in such an investigation is the limitation of variables. Ideally, the conditions to be met are that:

1. One has a series of compounds involving one anion and a wide variety of metal ions.

2. During the spectral determinations, the compounds should be in the vapor state, and dilute, so that the metal ion and its associated anions are unperturbed by "next nearest neighbor effects", and the temperature should be reasonably low, so that the contributions of vibronic transitions are minimized.

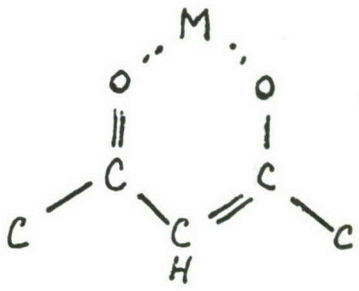
Unfortunately, no extensive group of inorganic compounds are volatile at room or lower temperatures. A compromise then would be to dissolve the compounds in an inert solvent and to study the spectra of dilute solutions. The compounds must, therefore, be soluble in an inert solvent such as hexane or iso-octane – an unusual property for inorganic compounds – and must absorb intensely, so that their spectra may be studied in extremely dilute solutions.

Success using this method is critically dependent on the ability to find a group of compounds that have these characteristics. Considerable study of available compounds eventually led to the selection of the metal acetylacetonates.

The parent acid, acetylacetone, "HA", has a simple spectrum, with a single essentially symmetrical, intense (e_m = 11,000) band at 271 mμ (1 mμ = 10A^{o} and e_m is the molar extinction coefficient). In general, all of the metal acetylacetonates have spectra which include components of the same order of intensity of absorption. They are sufficiently soluble

(about 10^{-5} moles per liter are needed for the spectral determinations) in saturated hydrocarbons such as iso-octane for the purposes of the study. Iso-octane was chosen as the solvent because it is non-polar, it is transparent down to 200 $m\mu$, it may be purchased in high purity, and may be further purified with minimum difficulty.

The acetylacetonate anion, which shall be abbreviated to "A", has a negative charge of one. Because of its chemical and electronic structure, however, it possesses two sites which can interact with metal ions. The representation



is unusually accurate because the group is planar. The associated metal ion (M) is, in effect, the sixth member of a hexagonal ring. Because the metal ion interacts with both oxygens of the anion, bis (acetylacetonates), MA_2, consist of a metal ion surrounded by four oxygens in a tetrahedral configuration or in a plane. In tris (acetylacetonates), MA_3, the six oxygens invariably appear to be in an octahedral configuration, and in tetra (acetylacetonates), MA_4, the eight oxygens occur in an archimedian antiprism configuration.

Because the intense components of the spectra of acetylacetone and the

metal acetylacetonates occur below 400 mμ, the present study was limited to the 200-400 mμ ultraviolet region of the spectrum.

EXPERIMENTAL

The metal acetylacetonates used in this study (Fig. 1) were prepared by previously described methods[1] or variants thereof. The anhydrous acetylacetonates of Be(II), Mg(II), Sc(III), Zr(IV), Th(IV), V(III), VO(II), Cr(III), Mn(II), Mn(III), Fe(III), Ru(III), Co(II), Co(III), Rh(III), Ir(III), Ni(II), Pd(II), Pt(II), Cu(II), Zn(II), Al(III), Ga(III), In(III), and Lu(III), were further purified by subliming at 150-200°C and 10^{-5}mm. Because of potential or observed instability above 150°C, the anhydrous acetylacetonates of Li(I), Na(I), K(I), Ca(II), Sr(II), Ba(II), Y(III), La(III), Cd(II), Hg(II), Tl(I), UO_2(II), and Ce(III) were used without further purification. They were held at 10^{-5}mm and room temperature for three days, however to ensure absence of water. The hydrates $MgA_2.2H_20$, $YA_3.8H_20$, $MnA_2.2H_20$, $CoA_2.2H_20$, $NiA_2.2H_20$, $ZnA_2.H_20$, and $U0_2A_2.H_20$, were prepared under conditions resulting in relatively large crystals, which were air dried and used with no further preparation. Carbon, hydrogen, and metal analyses were carried out on all specimens to ensure that their composition was as expected.

Spectrophotometric grade iso-octane was refluxed for 24 hours over metallic sodium in dried argon ambient before distilling. The first and last 10% of each batch of iso-octane distilled was rejected.

Spectral determinations were carried out using a Cary model 14 spectrophotometer calibrated against lines of the mercury spectrum. Glass stoppered cells with fused quartz windows and an effective path length of 10 mm were used.

RESULTS

The positions and, when available, the absolute intensities of components in the spectra of acetylacetone and metal acetylacetonates are presented in Table 1. The intensity of the bands is expressed as the molar extinction coefficient of the acetylacetonate groups present rather than of the entire complex. Thus, the value of e_m of the maximum of the 271 mμ band in the spectrum of FeA_3 is recorded as 10,300 rather than 30,900 which would be the value for the complex.

A very important factor in the determination of these spectra is the absence of moisture. Not only the intensity, but the very presence of some components is dependent on the presence or absence of even a trace of moisture. No quantitative data were collected on the amount of water which was necessary to produce certain changes, but it was observed that exposure to the open atmosphere during the preparation of solutions of some compounds, particularly those of LiA, NaA, KA, CaA_2, SrA_2, and BaA_2 produced significant changes in the spectra. On the other hand, the same solutions may be kept for six months in a desiccator without appreciable change in their spectra.

The spectra of the metal acetylacetonates were observed at one concentration level only (about 10^{-5} molar). This is important because it means that only those spectral components with absolute intensities (e_m) greater than about 1000 were detected.

DISCUSSION

I. Group Relationships

A. Classification of Spectra

The spectra of the metal acetylacetonates in the 200-400 mμ

region may be represented as in Figure 2. On the basis of complexity, the spectra may be divided into three groups:

1. Those which, when normalized with respect to the intensity of the band maximum are comparable to that of the undissociated parent acid, acetylacetone (HA). The spectrum consists of one essentially symmetrical band with its maximum at 271 mμ, and the distribution evident in Figure 3.

2. Those which consist of one major band (invariably at or above 283 mμ) and a shoulder at about 303 mμ or 315 mμ.

3. Those which consist of at least one major band and two or more secondary bands.

Hereafter, the spectra will be referred to as types 1, 2, or 3, depending on the above classification.

B. Types of Compounds Involved

Compounds with type 1 spectra include those of mono-, di-, tri- and tetravalent metal ions. Both paramagnetic and normally diamagnetic metal ions are represented. Some are anhydrous but others are hydrates. All of the hydrates considered have type 1 spectra, but a number of the corresponding anhydrous compounds have type 2 or 3 spectra. With one exception (ZrA_4), all anhydrous compounds with type 1 spectra involve metal ions of large ionic radius and/or low ionic charge.

Compounds with type 2 spectra are those of normally diamagnetic metal ions, e.g., Be^{++}, Al^{+++}, Th^{++++}, of small or moderate ionic radius and a valence greater than one. Compounds with type 3 spectra are those of transition metal ions of moderate ionic radius and a valence of at least 2.

C. Implication of Above Relationships

These data indicate that:

1. The electronic states of the anion are appreciably modified only as a consequence of association with certain metal ions. This is contrary to conclusions of previous investigators that all metal ions were influential, but that so many characteristics of the metal ions were involved that virtually no order was apparent in the spectra[2] or else that with increasing charge of the metal ion, there should be a successive shift of the major band to higher wavelengths[3], and to the effect noted in the spectra of metal ions. Assuming constant symmetry, as the ligand field of a metal ion increases in intensity, the components in the spectrum of the metal ion generally shift to a proportionate degree. It appears, however, that there is a threshold metal ion-anion bond intensity which must be exceeded before the electronic states of the anion are influenced by the metal ion.

2. The valence of the metal ion is not by itself a factor determining the ability of the metal ion to exert any influence on the anion's electronic states.

3. Neither the presence nor absence of unpaired d-electrons in the metal ion is a factor determining the ability of the metal ion to exert any influence on the anion's electronic states.

4. In those anhydrous compounds with type 2 or 3 spectra, the expansion of the coordination sphere to include one or more molecules of water is sufficient to destroy the metal ion-anion interaction.

5. The greater complexity of type 3 spectra (of transition metal compounds) as compared to those of type 2 (of non-transition metal compounds) suggests that one or more of three conditions exist in the

transition metal complexes:

a. As a consequence of π-bonding, the degeneracy of states of the anion is removed, making possible a greater number of transitions.

b. Transitions of electrons from the anion to unfilled metal ion orbitals are taking place.

c. Transitions from the metal ion's d-orbitals to unfilled orbitals of the anion are taking place.

D. Polarizability Effects

Although in general, the compounds of largest metal ions have type 1 spectra, and those of the smallest metal ions have type 2 or 3 spectra, there is no threshold ionic radius dividing the compounds with differing spectra. It has already been observed that there is no general classification of the spectra on the basis of the valence of the metal ions involved. If, however, one considers the values of Z/r (Z = ionic charge, r = ionic radius) for the metal ions (Table II), it is observed that when Z/r is less than 2.6, the spectrum of the acetylacetonate is of type 1. When Z/r is greater than 2.9, the spectrum is of type 2 or 3. When Z/r is between 2.7 and 2.9, the spectrum may be of type 1, 2, or 3. A number of compounds in this last group are reported to be dimorphic[4] or capable of existence as dimers or trimers[5].

No clear theoretical significance can be given to Z/r, but in a naive sense, it can be considered a measure of the polarizing ability of the metal ion. As such, the observed relations indicate that in those compounds with type 2 or 3 spectra, the supposedly non-bonding electrons

of the anion's oxygens are polarized along the metal ion-anion bond axis, and that in those compounds with type 1 spectra, the electronic states of the anion are at some equilibrium value from which they are dislodged only after the application of some minimum force.

II. Type 2 Spectra

Although the type 2 spectra (Fig. 4) are alike in having one major band and a strong shoulder at higher wavelengths, they can be further subdivided:

1. The spectra of the bis(acetylacetonates) – BeA_2, MgA_2, and ZnA_2 – are alike in having a major band at 292 ± 1 mμ and a shoulder at 303 ± 2 mμ . Observed variations might be attributable to differences in the relative contributions of two overlapping bands.

2. The spectra of the tris(acetylacetonates)– AlA_3, GaA_3, InA_3, YA_3, LaA_3, and LuA_3 – are alike in having one major band at 286 ± 1 mμ and a shoulder at 303 ± 2 mμ : The spectrum of ScA_3, with its major band at 299 mμ , is anomalous.

3. The only tetra(acetylacetonate) in this group, ThA_4, has its major band at 283 mμ and a shoulder at 315 mμ .

The most obvious variable among the compounds with type 2 spectra is the valence of the metal ion. The charge of the metal ion is often considered an important variable because one associates strong polarizing ability with high charge. Polarizing ability, however, is a function not only of the charge, but also of the radius of the metal ion. One would expect, therefore, a significant difference between the spectra of BeA_2 (r_{Be} = 0.3 A) and ZnA_2 (r_{Zn} = 0.7 A) if polarization of the anion's orbitals were involved

in the spectral changes.

Belford, Martell, and Calvin[3] attempted to explain the spectrum of the anion in metal acetylacetonates as the consequence of the coulombic interaction of metal ion and anion. Using simple LCAO-MO calculations, they determined the ground state and the excited state charge distribution of the acetylacetonate group and then assessed the effect of increasing the charge of the associated metal ion by one unit. From their calculations, they concluded that increasing the charge of the cation by one unit should cause the band in the spectrum of the anion to shift to higher wavelengths by 2800 cm^{-1}. It is evident from both the type 1 and 2 spectra that Belford, Martell, and Calvin's conclusions, and presumably their assumptions, appear to be incorrect. It might be pointed out, for instance, that the spectra of LiA, CaA_2, CeA_3, and ZrA_4 are all alike. When the metal ion does influence the position of the band in the spectrum of the anion, e.g., BeA_2, AlA_3, ThA_4, increasing charge of the metal ion is accompanied by a shift to lower wavelengths, not to higher wavelengths as they predicted.

At present there is no satisfactory model reported in the literature that explains, much less predicts, the observed shift of the major band in the spectrum of the anion with increasing charge of the associated metal ion. A possibility that does not appear to have been explored theoretically is that the spectral characteristics are in some way dependent upon the symmetry of the anions about the metal ion.

III. Type 3 Spectra

A. Subdivision of Spectra

Although the type 3 spectra are by far the most complex, they

may still be divided into subgroups on the basis of common spectral characteristics:

1. The bis(acetylacetonates), MA_2, have in common a major band at or above 292 mμ .

2. The tris(acetylacetonates), MA_3, have a major band at or below 272 mμ .

To the extent that the spectra of all of the bis(acetylacetonates) have a major high wavelength component (292 mμ or higher), though not necessarily a major low wavelength component, and those of the tris (accetylacetonates) have a major low wavelength component (272 mμ or lower), though generally not a major high wavelength component, these spectra resemble the type 2 spectra, in which there is a displacement toward lower wavelengths with increasing charge of the metal ion. It would appear, therefore, that the same fundamental factors are operative in determining the positions of these "characteristic" bands in both type 2 and 3 spectra.

B. Spectra of Bis(acetylacetonates)

On the basis of their spectral characteristics, the bis (acetylacetonates) may be divided into:

1. The non-planar compounds – NiA_2, VOA_2, and CoA_2, – whose spectra (Fig. 5) consist of a single major band (at 302, 297, and 292 mμ , resp.) and two or more secondary components.

2. The planar compounds – CuA_2, PdA_2, and PtA_2 – whose spectra (Fig. 6) consist of three or more major bands at about 325, 295, 250, 225, and/or 205 mμ , as well as secondary components.

The spectra of these compounds are beset with two complications not evident in the spectra of any other group considered. First, there is the vastly greater complexity of the spectra of the planar compounds as compared to those of the non-planar compounds. There is a very strong temptation to attribute the spectral differences to the structural differences. It is reported that NiA_2 is trimeric[5], with the nickel atoms being surrounded by a very distorted octahedron of oxygens. In dilute solutions, CoA_2 has a tetrahedral configuration[6], and VOA_2 has a rectangular pyramidal configuration with the acetylacetonate groups and vanadium atom almost in a plane[7]. The expected planar structure of CuA_2 has been verified[8], and it is reported that PdA_2 is isomorphous with CuA_2[9] The consistent planarity of tetracoordinate Pt(II) compounds leads one to expect that PtA_2 is also a planar molecule. The planarity of the CuA_2, PdA_2, and PtA_2, however, is the consequence of d-electron structure of the metal ions. This same d-electron structure of the metal ions, however, enhances the possibility that filled d-orbitals of the metal ions may interact with unfilled π-orbitals of the anion's oxygens. The existence of this condition is much less probable in the case of the Ni(II), Co(II), and VO(II) compounds. Thus, the differences in the complexities of the spectra of these transition metal tris (acetylacetonates) may stem from a more basic factor than molecular structure.

The second complication in the spectra of these compounds is in the range within which the "characteristic" band of these spectra occur. In every other group with type 2 or 3 spectra, a band occurs at or very near to a particular wavelength. For instance, the type 2 spectra of the non-transition metal bis (acetylacetonates) include the major band

at about 292 mμ , while those of the tris(acetylacetonates) include the major band at about 286 mμ , and those of the transition metal tris(acetylacetonates) include a major band at either 271 mμ or 260 mμ . The spectra of the transition metal bis (acetylacetonates), as do those of the non-transition metal bis (acetylacetonates), include a major band at a wavelength above that characteristic of the band in the spectra of the tris (acetylacetonates). In the spectra of four of these compounds – CoA_2, CuA_2, PtA_2, and VOA_2 – this band is only moderately displaced (at 292, 294, 295, and 297 mμ , resp.). This band occurs at 302 mμ in the spectrum of NiA_2, and at 328 mμ in the spectrum of PdA_2. The proposal that a structural factor was involved might have some plausibility if it were not that the band in the spectrum of PdA_2 is so unusual. As has been mentioned, PdA_2 has been found to be isomorphous with CuA_2 and is expected to be comparable in structure to PtA_2. In other characteristics, the spectra of PdA_2, PtA_2, and CuA_2 are alike, and completely unlike those of any other compounds. To further complicate the issue, the tetrahedral configuration of CoA_2 and the occurrence of the characteristic band at 292 mμ have exact parallels in the tetrahedral structure of BeA_2[10] and spectral band at this same wavelength. At present, no relationships present themselves to explain the complexities of the transition metal bis (acetylacetonates).

C. Spectra of the Tris (acetylacetonates)

On the basis of their spectral characteristics, the tris—(acetylacetonates) may be divided into two groups:

1. Those (Fig. 7) with a major band at about 271 mμ and secondary components at both higher and lower wavelengths – VA_3, CrA_3,

MnA_3, FeA_3 and RuA_3. Only in the spectrum of CrA_3 are other major bands found, and these occur at both higher and lower wavelengths.

2. Those (Fig. 8) with a major band at 260 mμ , another major band at lower wavelengths, and secondary components between 300 and 330 mμ – CoA_3, RhA_3, and IrA_3.

The compounds in the second group, in addition to being isoelectronic, though representing successive electronic shells, are similar in being diamagnetic, i.e., in a low spin condition. This would appear to be the only major non-spectral difference between CoA_3, RhA_3, and IrA_3 and the compounds in the first group, all of which contain unpaired electrons. Among the transition metal bis (acetylacetonates), PdA_2 and PtA_2 are also diamagnetic, which suggests that in compounds with this characteristic a fundamentally similar phenomenon is taking place. It may be that, as already suggested for the planar bis (acetylacetonates), the filled d-orbitals of the metal ion may interact with the unfilled π-orbitals of the anion's oxygens.

Among the spectra of the members of the first group – VA_3, CrA_3, MnA_3, FeA_3, and RuA_3 – the prominence of the band at 271 mμ suggests that this is the same transition in each compound. Barnum[11] carried out an extensive theoretical study of the spectra of the 3-d transition metal acetylacetonates and concluded that as a consequence of metal ion-anion bonding, the degeneracy of the anion's π-orbitals is removed, making possible as many as four transitions where only one was possible before. In addition, up to three charge transfer transitions – from metal ion orbitals to orbitals of the anion's oxygens – should then also be possible. Although

the results of the present experimental study do not coincide well with Barnum's calculations as far as those transitions arising solely from the anion's electronic states are concerned, there is reasonably good coincidence of the calculated and observed charge transfer bands – those secondary bands occurring above 300 mμ . This would suggest that those components of the spectra of the transition metal tris (acetylacetonates) above about 300 mμ are charge transfer bands and that the electronic states of the anion generated by the metal ion-anion $-\pi$ -bonding are essentially as predicted by Barnum. The poor coincidence of theoretical prediction and the present experimental evidence in the region below 300 mμ , however, suggests that the transitions involved in this region are not as proposed or predicted by Barnum[11] and earlier investigators[2,3] .

There is other evidence also that the transitions actually taking place are not those anticipated. In anions like the acetylacetonate ion, the electrons are basically of three types:

1. There are those that are involved in the principal, skeletal, or σ -bonds between atoms. The energy levels of the electrons involved in these bonds are such that transitions involving them occur in the far ultraviolet region of the spectrum – below 200 mμ .

2. There are those involving secondary, or π -bonds between atoms whose energy levels are such that transitions between them generally lie in the near ultraviolet or visible region.

3. There are those electrons that are not involved in any bonds between atoms, and the states of these electrons are designated n-states. Such electrons are found in atoms such as nitrogen, sulfur, and

oxygen. It is the transitions between the ground (σ, π, and n) and excited (corresponding starred) states of these electrons that are believed responsible for absorption in the acetylacetonate group, and in particular, three types of transitions appear to have been identified:

$$\pi \rightarrow \pi^*$$

$$n \rightarrow \pi^*$$

$$n \rightarrow \sigma^*$$

Of these transitions, only the $\pi \rightarrow \pi^*$ transition ordinarily has a high probability. Kasha[12] made a study of the manner in which electronic states involved in $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions are influenced by the environment of the ion, and concluded that if the solvent is changed from iso-octane, for instance, to alcohol, and then to water, the band due to a $n \rightarrow \pi^*$ transition will shift to lower wavelengths, while one due to a $\pi \rightarrow \pi^*$ will shift to higher wavelengths. There appears to be no evidence that bands due to $n \rightarrow \sigma^*$ transitions will be influenced, and besides, they have a very low probability. The response of the band in the spectrum of acetylacetone, HA, to solvent changes is consistent with that observed for $\pi \rightarrow \pi^*$ transitions, and this is probably the basis for the assumption that the same transition is involved in the spectra of metal acetylacetonates. As part of the present investigation of the spectra of anions, a study was made of the manner in which the spectra of the metal acetylacetonates were modified by methanol and by water. It was observed that the band in the type 1 spectra is always shifted toward higher wavelengths while the major bands in type 2 or 3 spectra are shifted toward lower wavelengths. From these observations, one would conclude that in those compounds with type 2 or 3 spectra the nature of the electronic states involved in transitions are

fundamentally different than in those compounds with type 1 spectra, and that they have $n \rightarrow \pi^*$ character. It will take further investigation to determine whether the character of the electronic states in HA and those metal acetylacetonates with type 1 spectra have been modified by metal ion-anion interaction or whether different electronic states are involved.

IV. Spectra and Solvents

Cotton and Holm[2] carried out the most extensive study of the spectra of metal acetylacetonates prior to the present one. Using chloroform, ethanol, or water as solvents, they established the spectra of 31 compounds and concluded that "λ_{max} of the main absorption band is a sensitive function of a number of parameters, the most important being probably cation charge, cation size, and its distortion effects on the ring π-system and the π-bonding ability of the metal ion. In most cases these effects are inseparable and lead to a near random placement of λ_{max} as can be seen from the data...." In contrast, the present study demonstrates that there is a high degree of order in the spectra of the metal acetylacetonates. The randomness of their results, and the reason their spectra and the spectra of these compounds as determined by other investigators[13] differ appreciably from the results of the present study is that the solvents they employed either interacted with the compounds directly, or else contained contaminants that do, a fact they did not recognize. Chloroform and ethanol or other alcohol have been the most generally used solvents. Both chloroform and ethanol in their uncontaminated form will interact with some metal acetylacetonates at least to form solvates[14] – a fair indication that the solvent may alter the spectrum of the compounds. Commercial absolute ethanol contains more

than 0.1% water initially, and picks up more water very quickly on exposure to the atmosphere. The solutions used in spectral studies contain about 10^{-5} moles per liter of the metal acetylacetonate. If the metal acetylacetonate has any tendency to interact with water, ethanol, or chloroform, the solvents commonly employed for these studies contain enough reactive materials to do the job. As already mentioned, this investigator has established the spectra of the metal acetylacetonates in anhydrous methanol and in water, as well as in other solvents. It was observed that the data presented by Cotton and Holm[12] and other investigators of the spectra of these compounds[13] can be duplicated using reactive solvents or "wet" solvents. In specially dried iso-octane, or other saturated hydrocarbons, such as cyclohexane, the spectra of the metal actylacetonates were found to be initially identical and unchanged after six months. When wet or reactive solvents were used, not only are the spectra often different initially, but changes in and/or a general deterioration of the spectra are apparent in hours.

CONCLUSIONS

Contrary to all previous work, this study has demonstrated that the spectrum of an anion does reflect, in an ordered manner, the character of the metal ion with which it is associated. This is represented schematically in Figure 9. In order of importance, the characteristics of the metal ions appear to be:

1. Polarizing ability. This was measured in terms of Z/r, where Z is the charge of the metal ion and r is its radius.

2. The presence or non-presence of d-orbital electrons, i.e.,

whether they are transition or non-transition metal ions.

3. Valence, i. e., whether two, three, or four anions are associated with each metal ion.

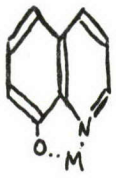
4. The strength of the metal ion-anion interaction. This was reflected in the planarity or non-planarity of bis(acetylacetonates) and in the paramagnetic or diamagnetic states of the metal ions in the tris (acetylacetonates).

It would seem that polarizing ability as measured here, the valence of the metal ions, the planarity or non-planarity of the molecules, and the magnetic condition of the metal ions are not fundamental enough properties to fully account for the differences in the character of the spectrum of the acetylacetonate ion in its compounds. The breakdown of the spectra into groups in which these factors appear important, however, provides the theoretician with substantial tools to be used in a consideration of the fundamental factors which determine the manner in which the electronic states of an anion is influenced by the metal ion with which it is associated.

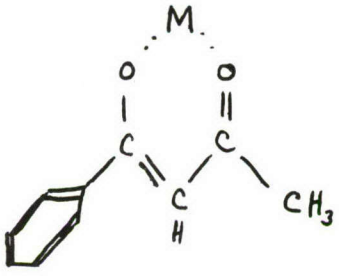
In the 70 or more years that the metal acetylacetonates have been investigated, the spectra of many of them have been reported several times, usually with significant differences among themselves and different from the spectra of these compounds in iso-octane or like solvents. Very few of these have any value as sources of information on the fundamental relation between physical and chemical properties of compounds because they include an uncontrolled variable whose influence on absorption spectra is neither appreciated nor understood. In the light of the present findings, all conclusions based on studies of the spectra which involve such reactive solvents or

contaminants must be carefully reconsidered.

The present investigation has shown that the electronic states of an anion are under certain circumstances influenced by the metal ion with which it is associated, and that a pattern of interrelationship of the spectral, physical and chemical properties exists. This has been demonstrated for a single anion but we do not know whether this is a unique situation, evidence of certain limiting circumstances, or a general situation. To try to clarify this point, the spectra of two other anions will be investigated. One of these will be the 8-quinolinate ion,



which is chemically quite different from the acetylacetonate ion. The other, the benzoylacetonate ion,



has the double virtue of being closely related to the acetylacetonate ion and of being potentially important in laser applications. Europium benzoylacetonate in solution lases readily, but it has been reported that this ability is solvent sensitive.

REFERENCES

1. Inorganic Syntheses, Vols. 1 thru VII; McGraw-Hill, New York, 1939 to 1964

2. Holm, R. H., Cotton, F. A., J. Amer. Chem. Soc. 80, 5658-5663 (1958)

3. Belford, R. L., Martell, A. E., Calvin, M., I. Inorg. Nucl. Chem. 2, 11-31 (1956)

4. Shibata, S., Bulk, Japanese Chemical Society 30, 842-4 (1957)
   Wold, A., Ph.D. Thesis, Polytechnic Institute of Brooklyn (1952)

5. Bullen, G. J., Mason, R., Pauling, P., Nature 189, 291-2 (1961)
   Graddon, D. P., Nature 190, 906-7 (1961)

6. Cotton, F. A., Soderberg, R. H., Inorganic Chemistry 3, 1-5 (1964)

7. Dodge, R. P., Templeton, D. H., Zalkin, A., J. Chem. Physics 35, 55-67 (1961)

8. Koyama, H., Saito, Y., Kuroya, H., J. Inst. Polytech. Osaka City Univ. 4, 43-49

9. Maki, A. H., McGarvey, B. P., J. Chem. Physics 29, 31 (1958)

10. Amirthalingam, V., Padmanabhan, V. M., Shankar, J., Acta. Cryst. 13, 201-4 (1960)

11. Barnum, D. W., J. Inorg. Nucl. Chem. 21, 221-237; 22, 183-191 (1961)

12. Kasha, M., Brealey, G. J., J. Amer. Chem. Soc. 77, 4462-8 (1955)

13. Sone, K., J. Amer. Chem. Soc. 75, 5207-5211 (1953)
   Kiss, A., Csaszar, J., Acta Chemica Hung. 13 (1/2) 49-57 (1956)

14. Steinbach, J. F., Burns, J. H., J. Amer. Chem. Soc. 80, 1839-1841 (1958)

TABLE I

Wavelength (mμ) and intensities of band maxima (as "A" molar extinction coefficient) of the spectra of acetylacetonates in the 200-400 mμ region. Shoulders are in parentheses.

Compound	max (mμ)	Cm
HA	270	11,000
LiA	271	
NaA	271	
KA	270	
TlA	271	
CaA_2	271	
SrA_2	271	
BaA_2	271	
CdA_2	271	
HgA_2	272	
PbA_2	271	
MnA_2	272	12,000
UO_2A_2	272	6,000
CeA_3	271	10,000
ZrA_4	273	
$MgA_2 \cdot 2H_2O$	270	10,000
$NiA_2 \cdot 2H_2O$	272	10,000
$CoA_2 \cdot 2H_2O$	271	7,000
$ZnA_2 \cdot H_2O$	271	11,000
$UO_2A_2 \cdot H_2O$	271	
$MnA_2 \cdot 2H_2O$	271	
$YA_3 \cdot 8H_2O$	271	

TABLE I (Cont'd)

Compound	max (mμ)	Cm	
BeA_2	293	16,000	
	(302)		
MgA_2	291	17,000	
	(304)		
ZnA_2	292	11,000	
	(302)		
CoA_2	292	8,000	
	(220)		
	(316)		
NiA_2	302	10,000	
	(217)		
	(273)		
	(284)		
	(316)		
VOA_2	297	8,000	
	207		
	(270)		
	(286)		
	(315)		
CuA_2	294, 297	14,000	
	245	11,000	
	202	11,000	
	(310)		
	(305)		
PdA_2	205	21,000	
	228	16,000	
	328	6,000	
	(253)		
	(317)		
PtA_2	224		
	249, 254	1.41	relative
	295	1.0	intensities
	(324)		
	(344)		
	(362)		
	(378)		
	(390)		
AlA_3	285	11,000	
	(302)		

TABLE I (Cont'd)

Compound	max (mμ)	Cm
GaA_3	287	10,000
	(303)	
InA_3	286	11,000
	(305)	
YA_3	285	
	(315)	
LaA_3	285	
	(315)	
LuA_3	285	
	(315)	
ScA_3	299	10,000
	305	
VA_3	272	7,000
	237	2,000
	347	2,000
CrA_3	324, 332, 337	10,000
	272	8,000
	256	8,000
	216	15,000
	(293)	
MnA_3	271	8,000
	317	3,000
FeA_3	271	10,000
	238	4,000
	351	1,000
RuA_3	271	6,000
	350	2,000
	(222)	
CoA_3	228	15,000
	260	15,000
	311, 317, 322, 325	3,000
RhA_3	204	18,000
	260	3,000
	311, 318	3,000
IrA_3	260	8,000
	306	4,000
	(317) (321)	
ThA_4	283	8,000
	(315)	

TABLE II

Valence, Radius, and Spectral Type

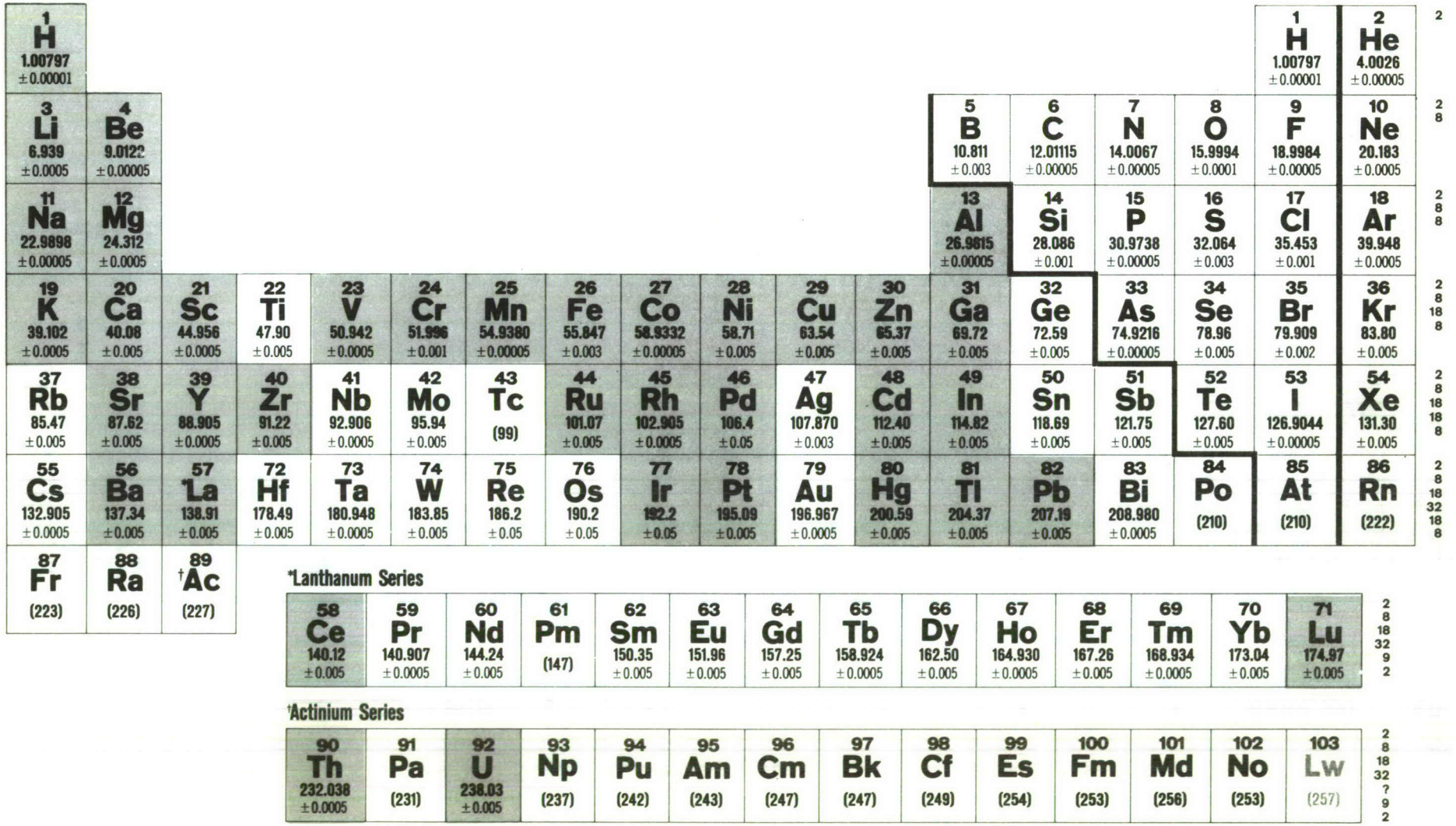
M	Z	r(Å)	Z/r	Spectral Type
Tl	1	1.40	0.7	1
K	1	1.33	0.8	1
Na	1	0.95	1.1	1
Ba	2	1.35	1.5	1
Li	1	0.60	1.7	1
Pb	2	1.20	1.7	1
Hg	2	1.10	1.8	1
Sr	2	1.13	1.8	1
Ca	2	0.99	2.0	1
Cd	2	0.97	2.1	1
Co	2	0.78	2.6	3
Ni	2	0.78	2.6	3
La	3	1.15	2.6	2
Zn	2	0.74	2.7	2
Ce	3	1.11	2.7	1
Mn	2	0.80	2.8	1
Cu	2	0.69	2.9	3
Mg	2	0.65	3.1	2
Y	3	0.93	3.2	2
Lu	3	0.93	3.2	2
In	3	0.81	3.7	2
Sc	3	0.81	3.7	2
Pt	2	0.52	3.8	3
Pd	2	0.50	4.0	3
V	3	0.51	4.1	3
Th	4	0.95	4.2	2

TABLE II (Cont'd)

M	Z	r(Å)	Z/r	Spectral Type
Cr	3	0.69	4.3	3
Ru	3	0.69	4.3	3
Fe	3	0.64	4.7	3
Co	3	0.63	4.8	3
Ga	3	0.62	4.8	2
Zr	4	0.80	5.0	1
Al	3	0.50	6.0	2
Be	2	0.31	6.5	2

Figure 1

Metal ions involved in the study are shaded.



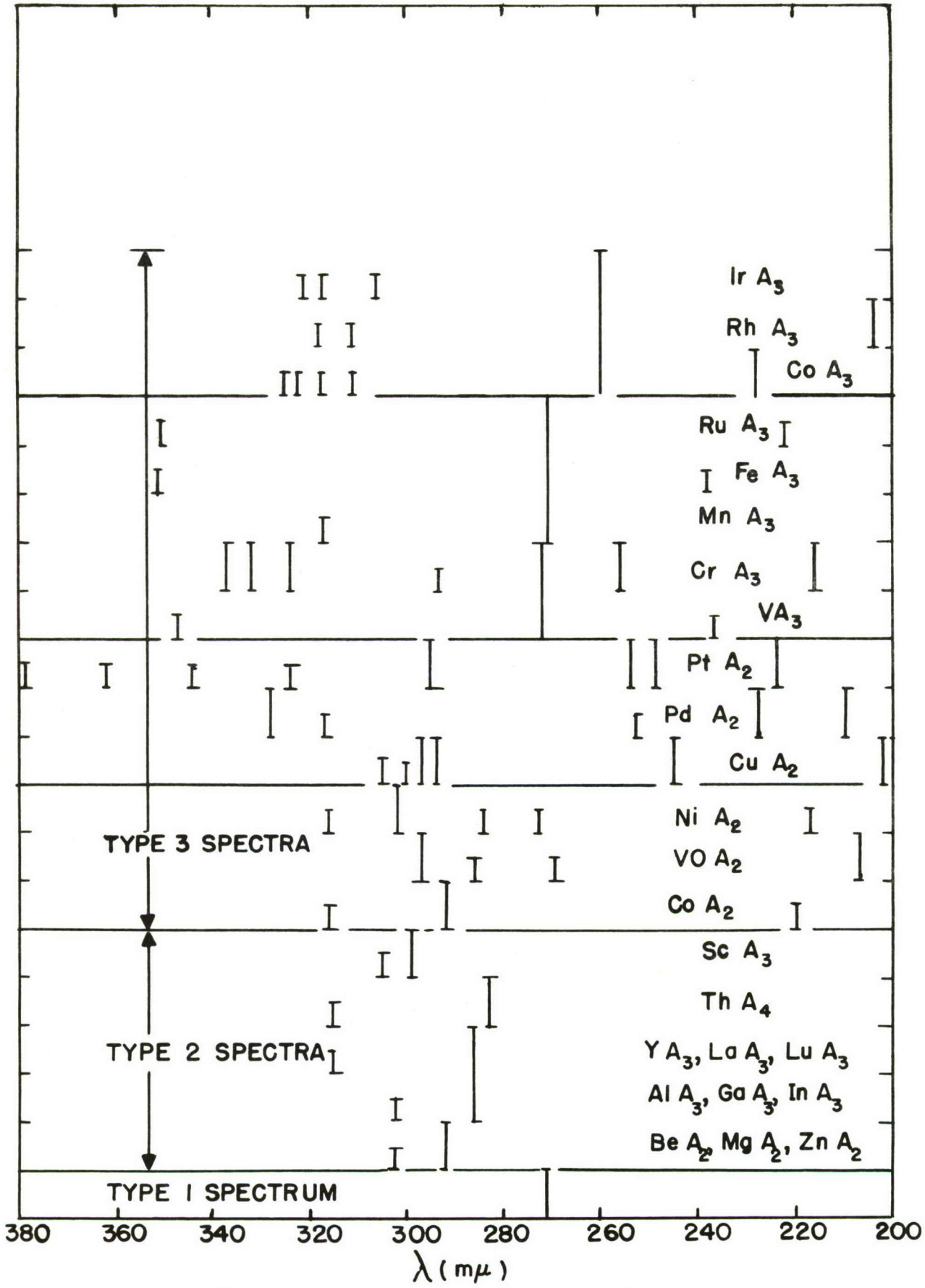


Figure 2 - Major (full height) and secondary (half height) components in metal acetylacetonate spectra.

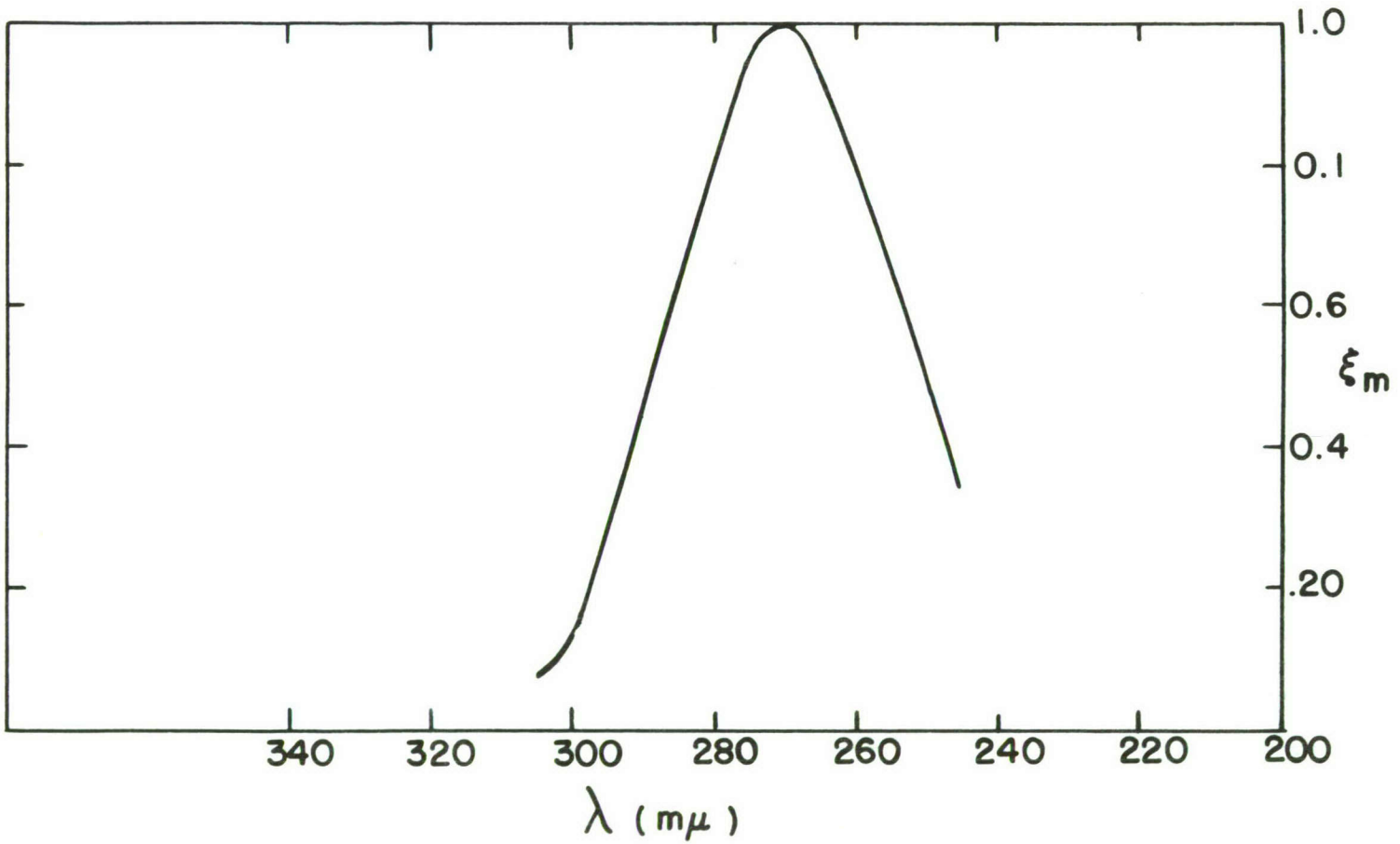


Figure 3 - Normalized band characteristic of Type 1 spectrum.

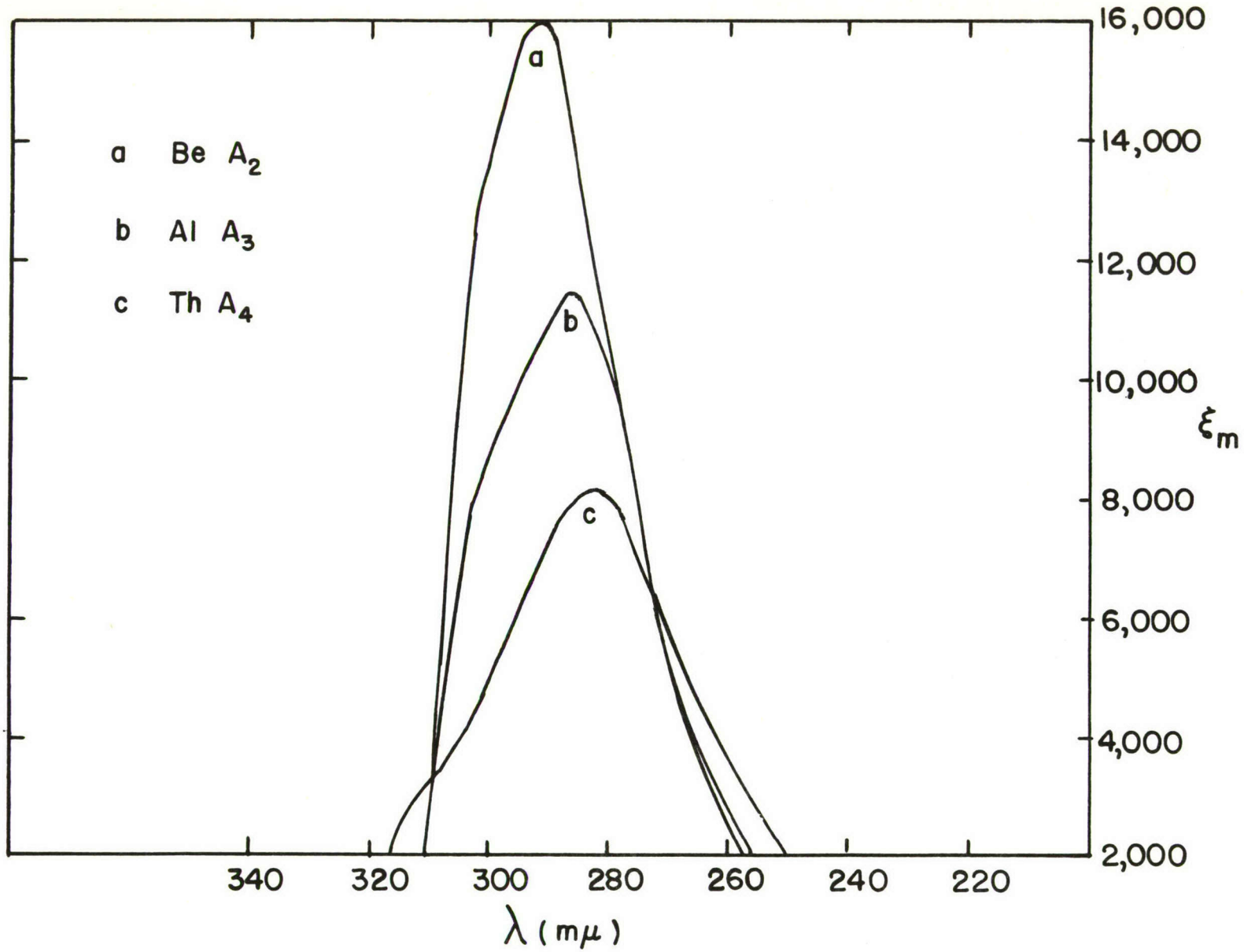


Figure 4 - Representative of Type 2 spectra.

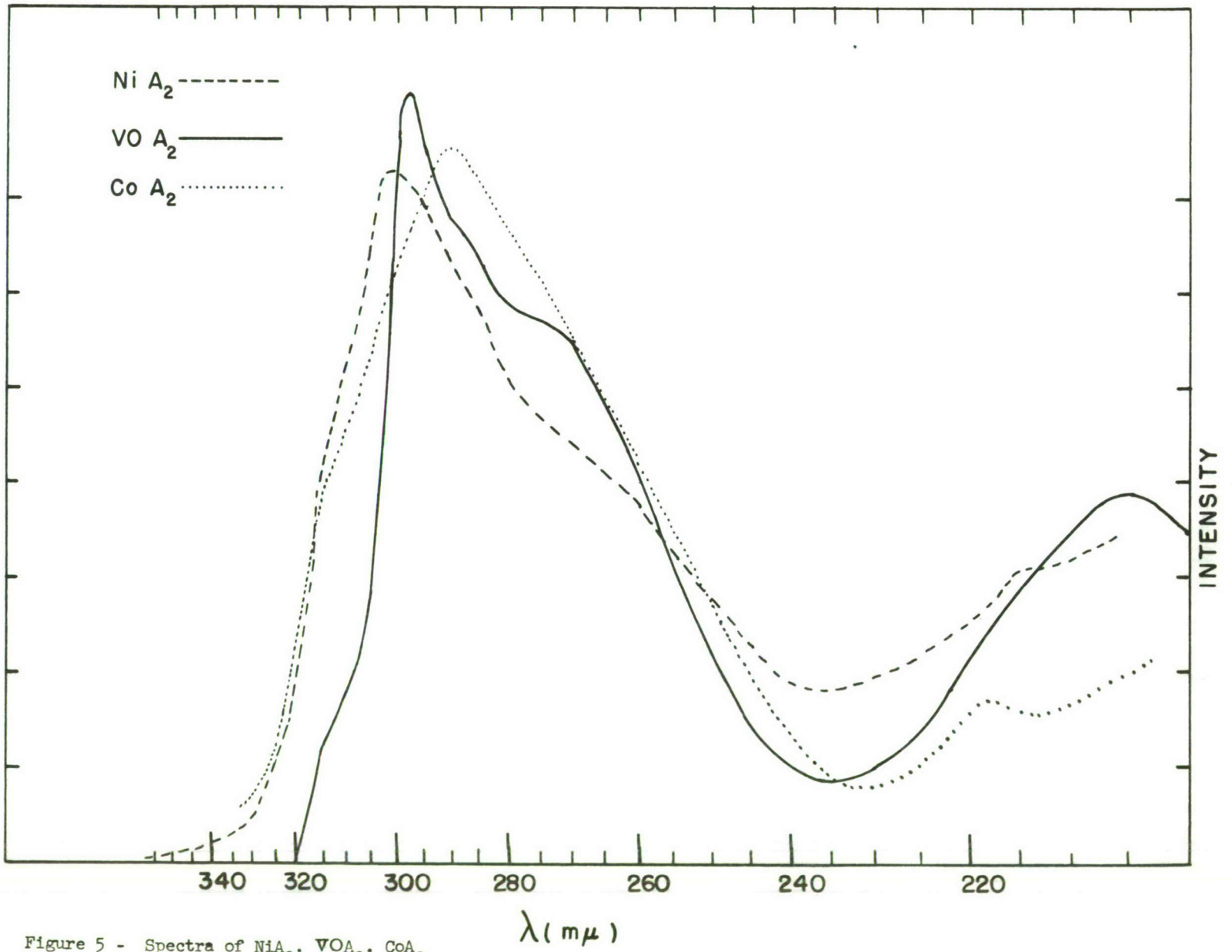


Figure 5 - Spectra of NiA_2, VOA_2, CoA_2

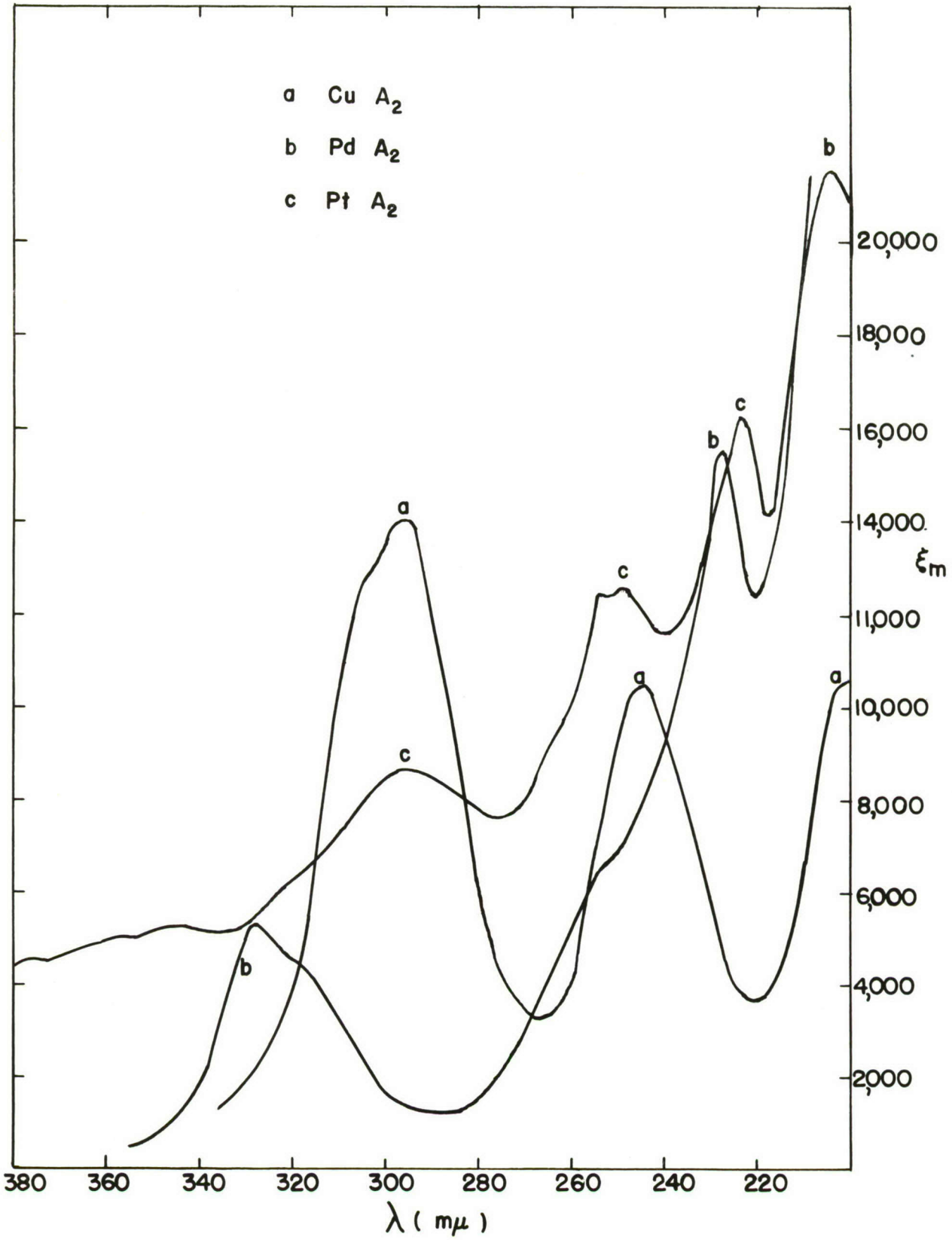


Figure 6 - Spectra of Cu A_2, Pd A_2, Pt A_2

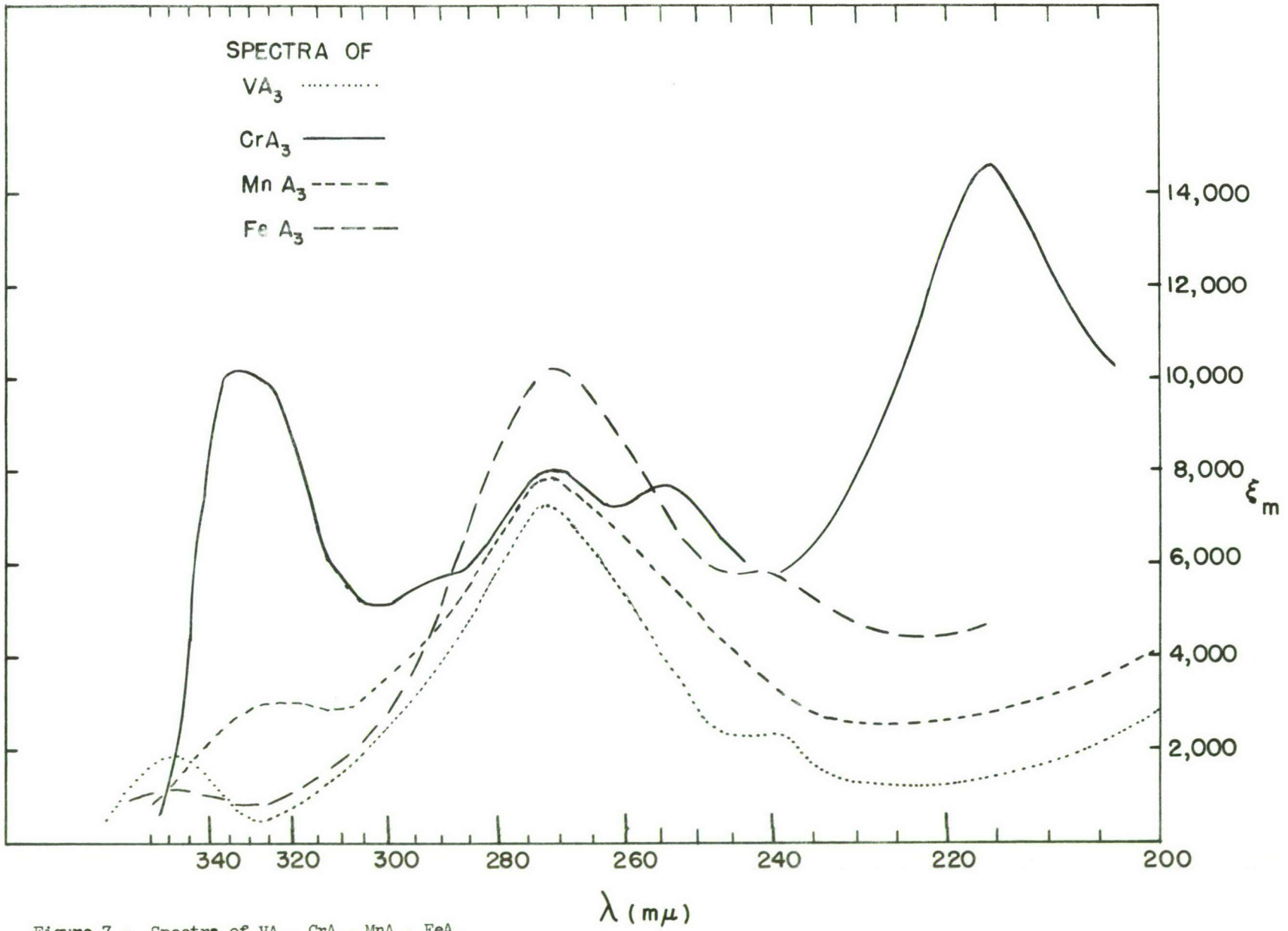


Figure 7 - Spectra of VA_3, CrA_3, MnA_3, FeA_3

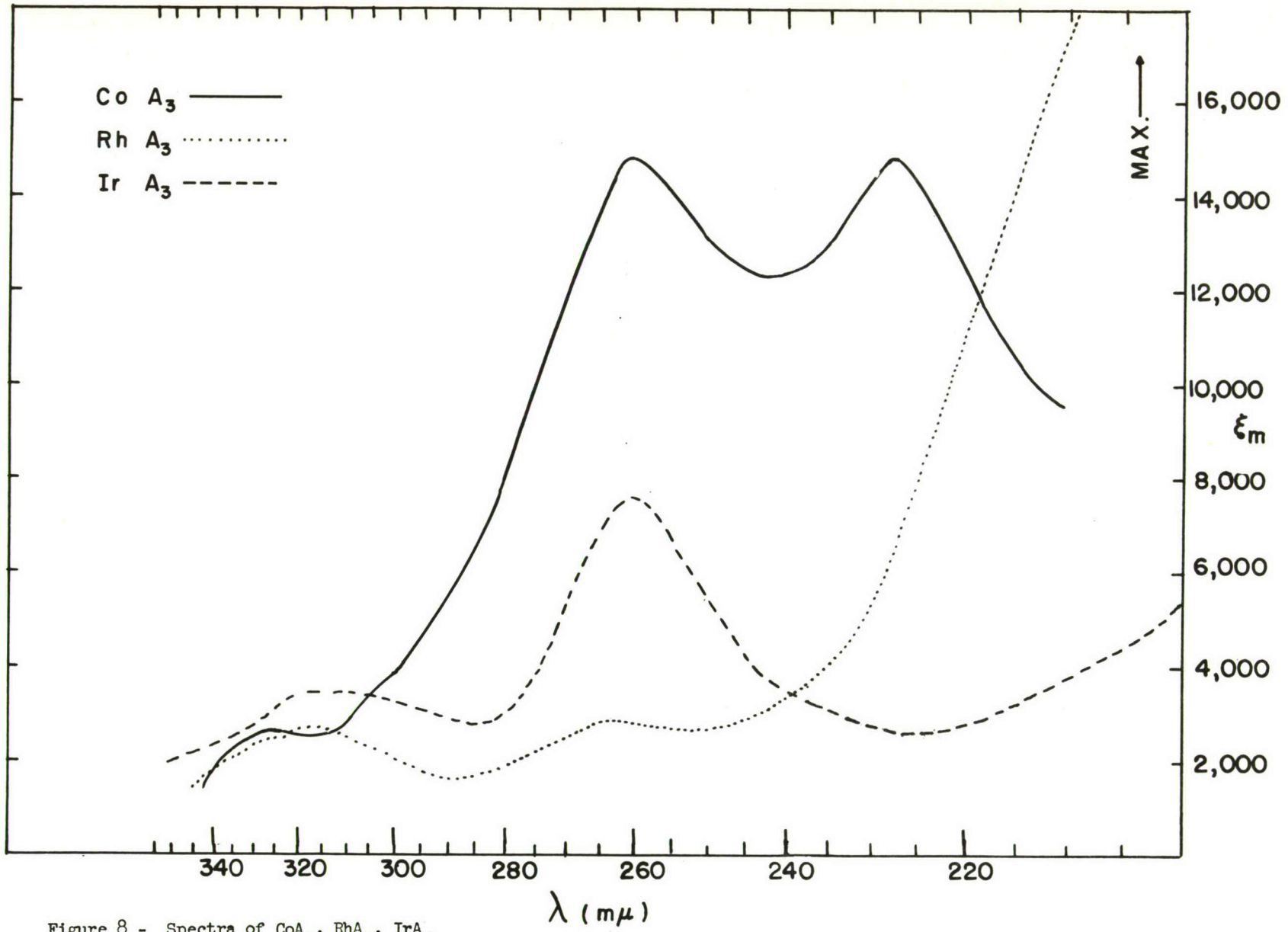


Figure 8 - Spectra of CoA_3, RhA_3, IrA_3

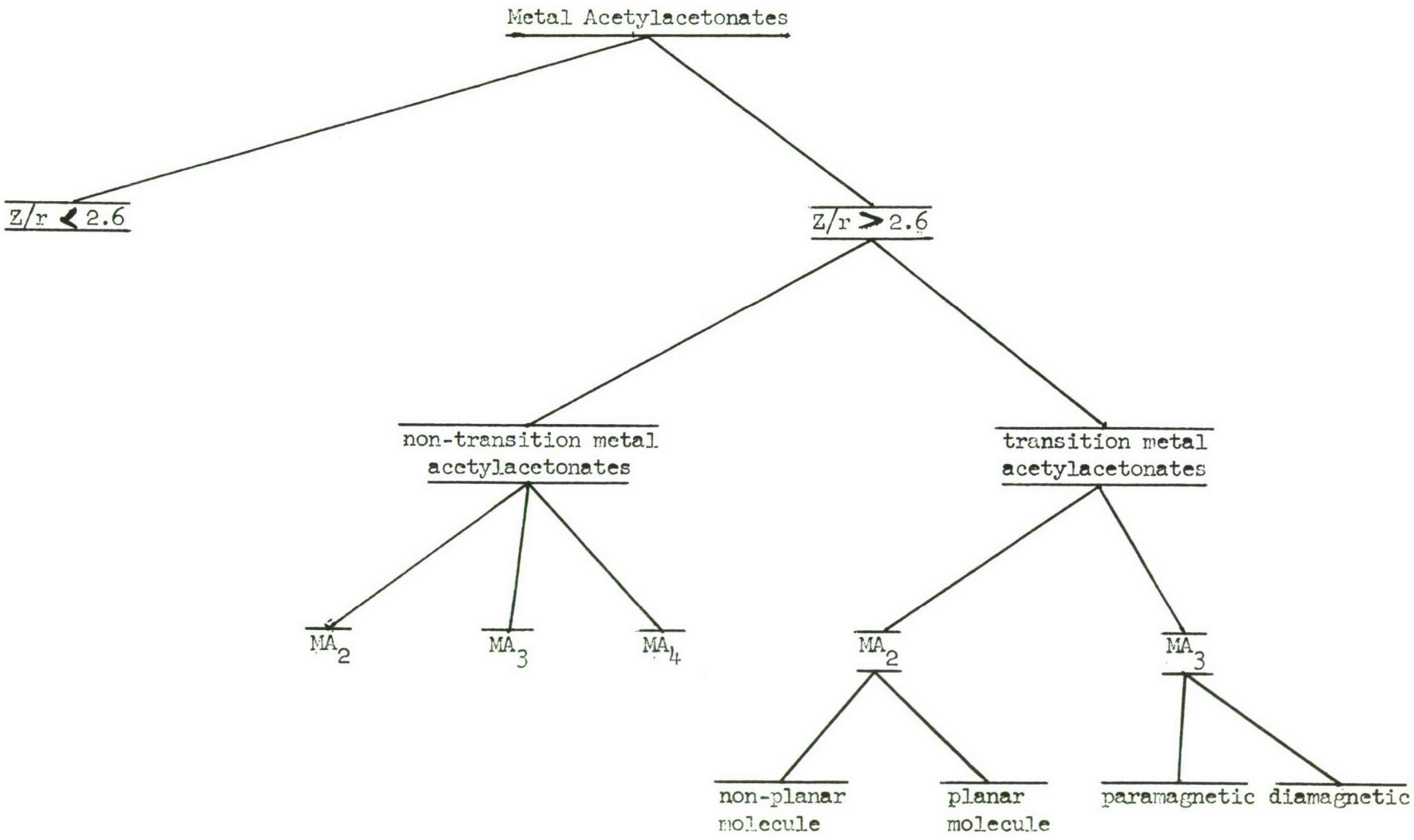


Figure 9 - Classification of metal acetylacetonate characteristics which parallel spectral groupings.

MEASUREMENTS OF RADIO STAR AND SATELLITE SCINTILLATIONS AT A SUB-AURORAL LATITUDE

Richard S. Allen, Jules Aarons, and Herbert Whitney

Air Force Cambridge Research Laboratories (OAR)

L. G. Hanscom Field, Bedford, Massachusetts

ABSTRACT

Observations of two radio stars, Cygnus A and Cassiopeia A, and of two satellites, Cosmos I and Transit 4A, have yielded data on lower and upper atmospheric irregularities. The frequencies studied have included 20 Mc, 40 Mc, and 54 Mc for satellite transmissions, and 30 Mc to 3000 Mc for radio star signals. The antennas used have ranged from a dipole to a 150 foot parabola. The irregularities in refractive index in the lower atmosphere produce amplitude fluctuations up to several times average with a fading rate of 3 to .5 per minute while ionospheric amplitude fluctuations can increase several decibels, with rates from 1 to 60 per minute. Lower atmosphere scintillation rates increase as the elevation angle increases and generally disappear above 5 degrees of elevation. During intense magnetic storms, the ionospheric scintillation rate is a function of wavelength in the UHF frequency range. Ionospheric scintillations are functions of both magnetic conditions and of the subionospheric latitudes of the irregularities; they are observed at zenith at the geomagnetic latitude of the Sagamore Hill Radio Observatory (54°). Using spaced receivers, heights of single irregularities have ranged from about 100 to 600 km and representative sizes from about 0.1 to 4 km. Irregularities constitute the prime factor in affecting the level, phase, and angle of arrival of signals propagated through the auroral regions.

MEASUREMENTS OF RADIO STAR AND SATELLITE SCINTILLATIONS AT A SUB-AURORAL LATITUDE

Introduction

The study of the atmosphere can be divided into two parts; one is cataloguing and understanding the regular variations of atmospheric parameters, and the other, cataloguing and understanding the superimposed irregularity structure. Depending upon the propagation mode to be studied, one or the other may be dominant. In high frequency long range communication, it is predominantly the mean electron density structure which deflects the wave for point to point transmission. In tropospheric and ionospheric scatter, it is the atmosphere's irregularity structure, its blobiness, that dominates. In auroral backscatter it is the patches and curtains of increased electron density which confuse radar echoes. In our study of radio star and satellite scintillations, we are emphasizing the irregularity structure of the atmosphere.

To select an analogue close to the subject of our scientific material, consider a small lens placed between an observing screen and a distant point source. The entire atmosphere has only a gross attenuation effect on the energy at the screen but the irregularity of refractive index, the lens, essentially focuses and defocuses the flux so that a distinctive pattern is formed at the screen. When there are many lenses, the resultant pattern is quite complex.

Using either cosmic radio sources or beacon satellites, we can study the irregularities in the ionosphere and to some extent those in the troposphere by examining the fluctuations received at a point on the earth's surface.

The combination of earth rotation, source motion, and ionospheric wind sweeps the irregularities through the ray path. The variation of refractive index in the irregularity produces variations in the phase front of the wave without attenuation of energy. As the wave travels onward, a phase and amplitude shadow pattern develops. Detecting this at the surface, one expects by analysis to determine the characteristics of the irregularities. In our measurements, only the amplitude component is recorded. Theory (Briggs and Parkins, 1963) indicates that for sufficiently small phase variation the scale of the shadow pattern is one to one with the scale of the irregularities.

The disturbance to clear propagation is evident. The fluctuations in amplitude from the lower atmosphere or from the ionosphere can drive the signal to values several times normal at peak and reduce it to background noise at null. This has been observed at frequencies from 7.5 to 9300 Mc/s. Fast fluctuations of the type encountered during auroral displays can produce intermittent loss of satellite telemetry or force an increase in redundancy to transmit the same information. Similarly, scintillation of phase can so change the angle of arrival that directional information is lost. During aurora, the phase deviations have produced inaccuracies in satellite tracking using phase interferometer techniques and in some cases have caused complete loss of signal. This is a serious problem when real time data must be obtained from satellites in polar orbit.

Observational Material/Equipment and Reduction of Data

The Sagamore Hill Radio Observatory of the Air Force Cambridge Research Laboratories is located at 43°N, 71°W, but its geomagnetic latitude is 54°N so that the ionosphere in the auroral region, up to the peak of auroral activity, 68°N, is above the radio horizon. From the point of view of

auroral effects, the site acts much like a similar observatory in Oslo, Norway. This site is, therefore, ideally suited for study of auroral activity, particularly for the study of solar interaction with the upper atmosphere.

Our probing instruments are both satellites and cosmic sources. A typical beacon satellite moves from horizon to horizon so fast that most ionospheric changes or motions may be ignored. The record displays the simultaneous behavior at a multitude of positions each examined at one or perhaps two frequencies. The in-range time is about 15 minutes and there is no further information until the satellite next arrives in view.

Using a cosmic radio star source, we may simultaneously observe at as many frequencies as we can instrument. Source motion due to earth rotation is relatively slow so that the ionosphere does change during a long observation. Periods of ionospheric disturbance can be examined in detail. Figure 1 is a projection of the intersection between the time varying ray path to a cosmic source and a constant ionospheric height of 400 km. Loci are shown for the intense sources, Cassiopeia A and Cygnus A. Cassiopeia A is circumpolar, traveling between 11° and 74° elevation while Cygnus A rises and sets in the auroral region. Both are in the northern sky at our latitudes so are seen through a disturbed ionosphere for at least some portion of their passage. Satellites, of course, can be seen within the entire circle enclosed by the radio horizon.

Satellite Equipment and Typical Record

Observations of the satellite Transit 4A on 54 Mc/s have been analyzed for the period July 61 to December 63. The receiving equipment consisted of dipole antennas, direct power receivers with time constant 0.1 second and chart recorders calibrated in power. Each pass of the satellite within range

was recorded. The index of scintillation activity was based on the average signal level at a Faraday peak. When signal fluctuations fell to the cosmic noise level, the index was defined 100%. Latitude and longitude under the intersection of the ray path and the ionosphere at 400 km was determined for each reading at the Faraday peak.

As an example of both the geometry and the type of signal observed, Figure 2 illustrates the record of amplitude fluctuations for a polar satellite transmitting at 54 Mc/s. For this pass, the satellite moved from a quiet ionosphere with little or no irregularities in the south into a disturbed auroral region where there are intense scintillations caused by strong irregularities. The small peaks and nulls are Faraday rotations of the signal but the heavy black lines are scintillations. Total electron content along the line of sight to the satellite can be derived from the Faraday nulls and the scintillations permit the study of ionospheric disturbance mechanisms.

Cosmic Radio Source Equipment and Typical Record

Observations of the radio source Cassiopeia A and Cygnus A have been analyzed for the period October 1961 to January 1964. The 84 foot equatorially mounted antenna was used simultaneously with Dicke switched radiometers at 63, 113, 228, 400, 120, amd 3000 Mc/s. By 1962, a program had evolved whereby Cassiopeia A and Cygnus A were monitored on alternate half hours for four days, twice a month. When Cygnus A was below the horizon, or when other observing programs were not being conducted, then Cassiopeia A was tracked continuously, weather permitting. Since June 1963, the 150 foot altitude-azimuth antenna has been used at a fixed elevation of 12.5° to make daily observations of Cassiopeia A at 28.9, 63, 113, and 228 Mc/s simultaneously.

The chart records are essentially linear in power. By pointing the antenna at reference cosmic noise backgrounds when switching between sources, the average

source power was determined for each observation. The scintillation index of the power variation was defined in terms of the peak power fluctuations as:

$$S = \frac{P_{max} - P_{min}}{P_{max} + P_{min}}$$

This is consistent with the scintillation index measured from the satellite records.

A rate of fluctuation was determined by counting peaks. For consistency an arbitrary rule prescribed counting a fluctuation peak only if it was at least one third the preceeding peak.

Figure 3 is a comparison record taken near 1.5° elevation. The top records compare both the 84 foot and the 150 foot antennas at about 800 foot separation and also crossed linear polarization at the same antenna. No differences in scintillation are seen at this separation, nor at this frequency is there any polarization scintillation. The lower record shows that an extended source such as Cygnus X (angular size $>2^{\circ}$) does not scintillate in the presence of irregularities of small angular extent, while a smaller angular source, such as Cygnus A ($\sim 1.5'$) scintillates heavily.

Statistical Results

The scintillation indices were separated into groups by using the three hour index, K_{Fr}, of the magnetic observatory, Fredericksburg, Virginia. Increasing magnetic fluctuations of the earth's field are reflected by an increasing quasi-logrithmic index scaled from K = 0 to K = 9. While it has long been noted that magnetic K index was closely associated with visual changes in solar activity, it has recently been shown that it is related

to the velocity of the solar wind.

Figure 4 shows the latitude distribution of satellite scintillations for the period July 1961 to July 1962. Data has been sorted to a north south swath $\pm 8^{\circ}$ wide for quiet magnetic intervals (K = 0,1,2,3) and for disturbed magnetic intervals ($K \geq 4$). Note the greater index of scintillation at low angles in the north as compared to that in the south. It is also clearly seen that when the magnetic field is being disturbed the irregularity structure extends toward the south in a manner similar to the aurora's movement toward the equator during magnetic storms.

Using a second spearation by geographic sectors, we were able to show control of the irregularities by the local magnetic dip field. For each 45° quadrant, the angle of elevation of the samples remains the same. The mean scintillation index in the northwest sector is greater than that of the northeast. Figure 5 illustrates this difference. Since the magnetic declination of our site is about 16°W, our radio horizon extends further into the disturbance region in the northwest sector. The four southern sectors have been averaged together. To a first approximation, we can separate elevation angle effect and latitude effect by dividing the north-south variation by the east-west variation.

Radio star observations by themselves, cannot easily provide either latitude or elevation angle dependence. Because of their constant celestial path (Figure 1), source elevation angle is always associated with a sub-ionospheric latitude. If several sources are used then the effect of source angular size further modifies the data. The control of the irregularity distribution by the magnetic dip field should make the observations assymetric and indeed our Cygnus rise data does show a systematic qualitative difference from Cygnus set data.

The diurnal behavior of the 54 Mc/s scintillations of Transit 4A during 1961 - 1962 is shown in Figure 6, plotted against sub-ionospheric (400 km) latitude. Just after sunset, there are generally low indices to the south

of the station and a sharp increase in the indices to the north. During the night the depth of scintillation increased about two-fold for the region 30° to 45° geographic latitude. In the auroral region to the north, there is not much night time increase observed. One probable reason is that the auroral zone ionosphere is so disturbed that the signal is scintillating completely most of the time. Predawn observations show that during this time period the irregularities are more diffusely spread in latitude.

While the latitude effects of the diurnal behavior of the irregularity regions can be examined from the satellite records, the radio star records are reduced for a detailed diurnal study at nearly constant latitude and elevation angle. Seasonal effects cannot, however, be separated from diurnal variation. Figure 7 shows the diurnal behavior of Cassiopeia A at upper culmination for low magnetic activity. The rate of scintillation peaks later than the amplitude of scintillation. During periods of increased magnetic activity, the time of occurrence of the fastest rates moved toward sunset but the mean rates did not increase. The mean scintillation index is greater during increased magnetic activity but the general diurnal behavior is similar. Examination of Cassiopeia A records at lower culmination and for Cygnus A at transit show that the diurnal behavior has a weak variation with geomagnetic latitude. The strong diurnal behavior of scintillations is still one of the principle unexplained characteristics of the ionospheric irregularity structure.

Radio Star Scintillation Study

Frequency Dependence of Amplitude

It has long been known that the amplitude of scintillations decreased with increasing radio frequency. From simple considerations of weak diffraction by well behaved irregularities, an inverse square law dependence would be

expected and, of course, this was observed at UHF. In the HF region, the amplitude of scintillation is often 100%, in fact, positive peaks of radio star signals are often much greater than twice the normal level. These are clear indications that the rules of weak scattering are being violated. Recently Briggs and Parkin (1963) applied diffraction analysis to anisotropic irregularities with a two dimensional space distribution. Considering phase deviations of less than 0.7 radians at the irregularities, they derive an inverse linear dependence for very small irregularities (far field conditions) and inverse square law for very large irregularities (near field conditions).

A comparison of our arbitrary scintillation index at different frequencies can only be done with caution. An examination has shown that the ratio of the measured peak scintillation to the computed value of the rms scintillation increased with magnetic activity but that for a given record the ratio was essentially the same for each frequency. The factor relating our arbitrary index to the rms index, therefore, cancels to first order when the frequency dependence of simultaneous records is computed. Bearing this in mind, the frequency dependence for June 1963 to January 1964 data taken at 12.5 degrees elevation with the 150 foot antenna is shown in Figure 8. The data is plotted for quiet (K_F = 0,1) magnetic conditions. The mean frequency dependence in the region 63 to 228 Mc is between linear and inverse square law. The mean frequency dependence between 30 and 63 Mc/s is less than inverse linear, contrary to the assumption of small phase deviation. It is not unusual during intense magnetic disturbance to find that the amplitude of scintillation decreases with decreasing frequency even in the higher HF region. This is most probably the result of both intense phase deviation and multiple scattering. At low elevation angles, the fluctuations have disappeared completely for extended period during intense disturbances.

When the phase deviations in the irregularities are weak, then the shadow pattern on the ground is in direct correspondence to the irregularity pattern in the ionosphere. Our multi-frequency observations have shown that this is not generally true for the usual HF satellite beacon frequencies. We have, therefore, only analyzed the rate of fluctuations of radio star observations at higher frequencies. The rate of fluctuations is an indication of both the changing geometry of the ray path and the velocity of the irregularities. Figure 9 shows the systematic increase in rate of scintillation with increase in magnetic activity (K_F). The data includes all Cassiopeia A observations during 1961 and 1962. Mean rates vary from about one per minute when $K_F = 0$ to about 4 per minute when $K_F = 5$. Rates of 12 or greater were often noted during visible aurora but the lack of sufficient data prevents statistical analysis of highly disturbed periods.

Height and Size of Irregularities

Although many efforts have been made to determine the height of the irregularity region from radio star observations, the varying latitude distribution casts doubts on those methods involving statistical comparison of records at varying elevation angle. If simultaneous reception of a satellite signal can be received at several widely spaced stations, then a geometrical determination can be made of the height of individual irregularities. Under the assumption that the irregularity is weak, localized and moving at much less than the satellite velocity, the height can be calculated from the time differences of sharp features in the diffraction pattern at the ground.

Using this technique, fast recordings of twin stations at Sagamore Hill have been analyzed. Typical results are shown in Figure 10. The number of points for which clear time differences could be measured is plotted against the time delay. The minimum, maximum, and most frequent

height of the irregularities is indicated. This range of heights is consistent with the results of Frihagen (1963).

Under the assumption that some of the irregularities are discrete blobs and are responsible for the recorded signal showing a periodic nature, i.e. null-to-peak-to-null, an estimate of the size of the irregularity can be made from the time between nulls. Knowing the height and velocity of the satellite and the height of the irregularity, there is a direct proportion between the distance traveled by the satellite and the extent of the irregularity along the satellite path. For those cases in the previous example where the data exhibited a definite period, the time between nulls was measured, converted into distance or size and the distributions are plotted in Figure 11. Considering the geometry at our site these indicate the scale across the small dimension of field aligned irregularities.

In addition to the fine scale irregularities so typical of both radio star and satellite observations, there is often evidence that the ionosphere is strongly disturbed over large localized regions. While fast switching between radio stars at different latitudes has suggested this possibility, it is best examined by multi-station satellite observations. For a single observatory, the satellite record shows a low fluctuation level followed by a region of intense scintillations and subsequently a return of a low level. The extent of such a region can be measured by time analysis of records from a number of observatories. In cooperation with several stations in Europe, cloud sizes were measured using the transmissions of Sputnik III. Results are displayed in Figure 12 where it is seen that the extent was roughly 200 by 400 km. Analysis also indicated that the cloud thickness could have extended from 250 km to at least the height of the satellite, 1000 km.

Tropospheric Scintillation

A radio wave also undergoes deviation by irregularities in the lower atmosphere. As with ionospheric scintillations, the variations of atmospheric refractive index produce a shadow pattern of peaks and nulls. These are clearly differentiated from concurrent ionospheric scintillations by examination of simultaneous frequency records as in Figure 13. For most of our data, ionospheric scintillations decrease to negligible amplitude above 400 Mc/s. The atmospheric scintillations in the upper microwave region have peaks several times the mean values of the source. The correlation between peaks and nulls at 1200 and 2980 Mc/s ranged from 0.4 to 0.7, indicating weak frequency dependence of the scintillation mechanism. Analysis of Cygnus rises and sets during the period 1961 - 1962 (Figure 14) shows that the mean rate of fluctuation increases as elevation angle increases, while the mean index decreases sharply. The day to day variations are considerable but show no correlation with obvious gross weather characteristics.

Discussion

The future military implication of any basic research problem is never easy to forecast but in this case it is not completely unforeseen. The large number of satellites traversing the auroral scintillation region points out the more obvious problems in signal degradation for surveillance or communication systems. Equally important in the years ahead are two probable applications. One is the use of scintillation data to aid in understanding the basic physics of the upper atmosphere; the second is the use of ground observations of the behavior of the irregularity regions as sensitive indicators of solar-terrestrial relationships, and perhaps even of man-made events in space.

Satellite studies have shown that northern latitude scintillations are very closely associated with the auroral zone, both in latitude distribution and in similar response to solar-geomagnetic activity. A few days after a major x-ray, optical, or radio wave outburst is observed on the sun, both the scintillation and the aurora may increase greatly in activity. Space detectors have shown that after some solar outbursts, the intensity and velocity of the solar particle flux in the vicinity of the earth is increased several fold and that these particles interact in some presently unknown way with the trapped radiation belts contained within the earth's magnetic field. The role of the radiation belts as a conveyer of this solar energy to the upper atmosphere is now beginning to be understood.

O'Brien (1964) has proposed a phenomenological model of these regions which describes the main behavioral features. There is always some particle precipitation from the trapped particle regions, the greatest flux tending to occur under the horns of the outer belt in the auroral zone. During a disturbed period, the average flux of precipitated electrons with energy greater than 40 KEV may increase by a factor of 10^3 when middle latitudes are compared to the auroral zone. Particle counts at various pitch angles have shown that the belt is not dumping its available particles, instead, there is actually an increase in the particle flux. This increase is primarily an increase in particles which penetrate to ionization heights within the auroral region. The precipitation flux can be either extensive in latitude and longitude zones i.e. of the order of several thousand kilometers or intensive in either coordinate, forming zones a few kilometers or less.

The general characteristics of the precipitated electron flux (McDiarmid et al, 1963) fit the observations of both the geographical and geomagnetic storm characteristics of scintillation. Moderately strong scintillations are always observed in the auroral zone (Little et al, 1963; Basu et al, 1963).

During disturbed periods evidenced by increasing magnetic activity (high magnetic index K), the increase in flux of precipitated particles corresponds to an increase in scintillation activity. The regions of scintillation irregularities move away from the auroral zone, just as visual auroral forms move towards equatorial latitudes. The increase in degree of disturbance is evidenced by an increase in depth of scintillation, particularly at the higher radio frequencies for auroral ray paths. Both flux and scintillation measurements have observed the presence of irregularity regions which at some times are very wide in geographic extent, as in diffuse aurorae, and at other times are very narrow, as in sharp auroral form.

The mechanism producing the scintillation irregularities at heights near 200 to 400 km need not be solely the precipitated electrons with energies greater than 40 kev measured near 1000 km, just as there is always the possibility for a third mechanism to produce the auroral reactions near 100 km. It does seem clear, however, that a basic inter-relationship has been established. It might be assumed that a more copious supply of lower energy particles is the coupling device between the solar disturbance and the ionosphere and that they are controlled in the same (unknown) manner as the 40 kev electrons.

The observational feature of scintillations that does not fit the precipitation observations is that of diurnal variation. Both satellite and radio star scintillation show distinct diurnal maxima near local magnetic midnight under all ionospheric conditions. O'Brien (1964) found little diurnal variation in the flux of precipitated electrons within the auroral region; he does not indicate diurnal variation at lower latitudes.

It is known, however, that the density of neutral particles in the upper atmosphere has a diurnal variation. In addition, the temperature of the constituent particles varies as a function of latitude, showing striking differences in the F region at sub-auroral latitudes (Delgarno, 1964).

Future Studies

The negative effects of the phase and amplitude fluctuations are considerable. A cataloguing of the effects is still necessary since there is not yet a basic understanding of their cause. This cataloguing should include those items of interest both to the research scientist and to the system designer, such as (1) correlation with upper atmospheric winds or storms, (2) latitude dependence, (3) peculiarities of the auroral and equatorial zone, (4) diurnal dependence, and (5) control by solar activity.

Once the properties of the scintillations are known, particularly the response to solar activity, then the negative effects can be avoided even if they are not completely understood. Operating frequencies then can be chosen by latitudes and stages of the sunspot cycle. Scintillations will no longer be a military problem, just part of an obstacle course. But the final study will provide information for using the irregularity structure for military missions, as a positive propagation tool similar to ionospheric and tropospheric scatter. An obvious example is the use of flat sheets of field aligned ionization in the equatorial regions as propagation ducts for transhorizon HF communication. In that way, we expect that scintillation studies will yield design information for communication systems.

References

Basu, Sunanda, R.S. Allen, and J. Aarons, "A Detailed Study of a Brief Period of Radio Star and Satellite Scintillations", J. Atmos. Terr. Phys., 1964 (in press).

Briggs, B.H., I.A. Parkin, "On the Variation of Radio Star and Satellite Scintillation with Zenith Angle", J. Atmos. Terr. Phys., 25, 339, 1963.

Dalgarno, A., "Corpuscular Radiation in the Upper Atmosphere", Ann de Geophysique, 20, 65, 1964.

Frihagen, J., "Fine Scale Ionospheric Structure by Satellite Techniques", in Radio Astronomical and Satellite Studies of the Atmosphere, J. Aarons, Editor, North-Holland Publishing Company, Amsterdam, 1963.

Little, C.G., G.C. Reid, E. Stiltner, and R.P. Merritt, "An Experimental Investigation of the Scintillation of Radio Stars Observed at Frequencies of 223 and 456 Megacycles per Second from a Location Close to the Auroral Zone", J. Geophys. Res., 65, 1763, 1962.

McDiarmid, I.B., J.R. Burrows, E.E. Budzinski, and M.D. Wilson, "Some Average Properties of the Outer Radiation Zone at 1000 KM", Candian Jour. of Phys., 41, 2064, 1963.

O'Brien, B.J., "High Latitude Geophysical Studies", Jour. Geophys. Res. 69, 13, 1964.

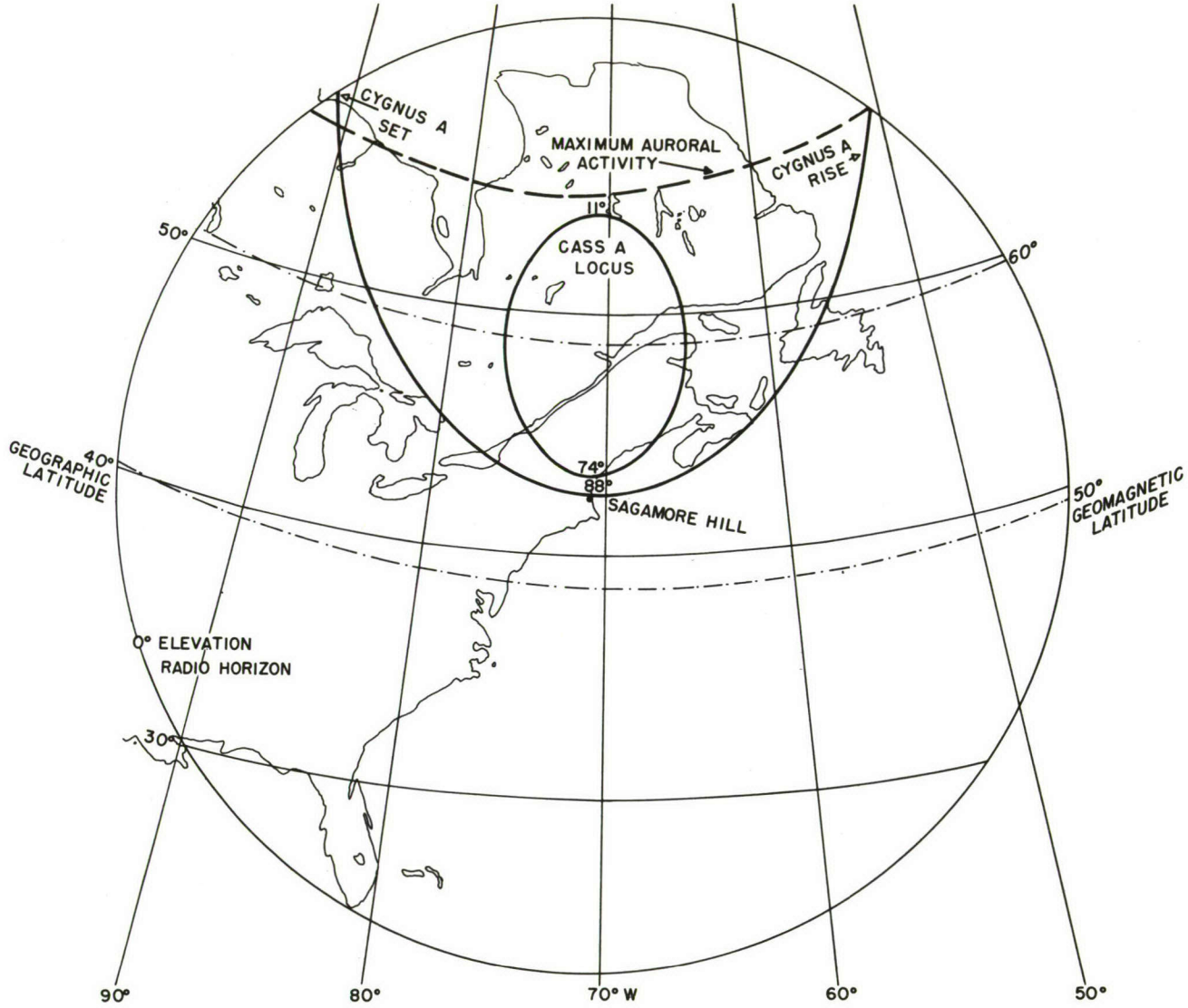
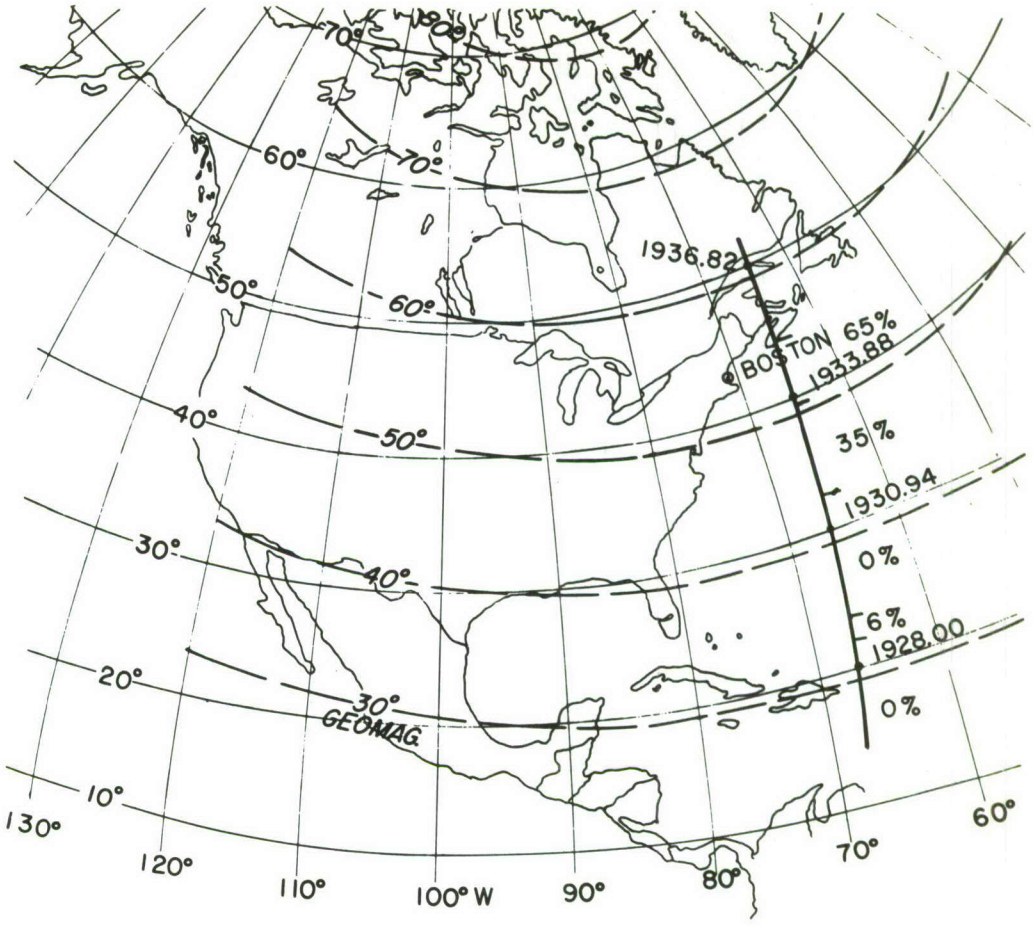
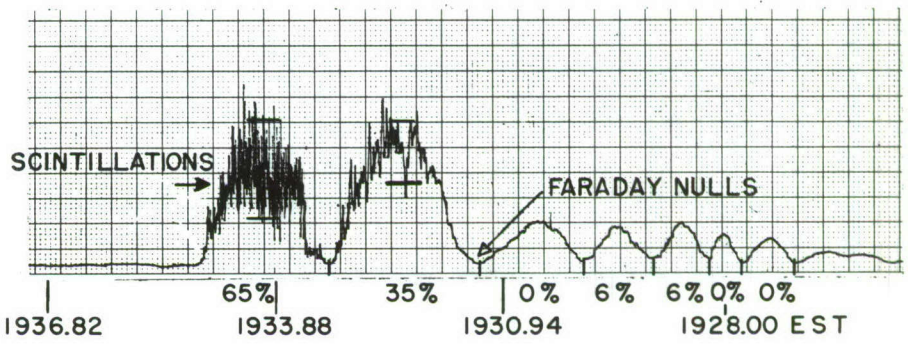


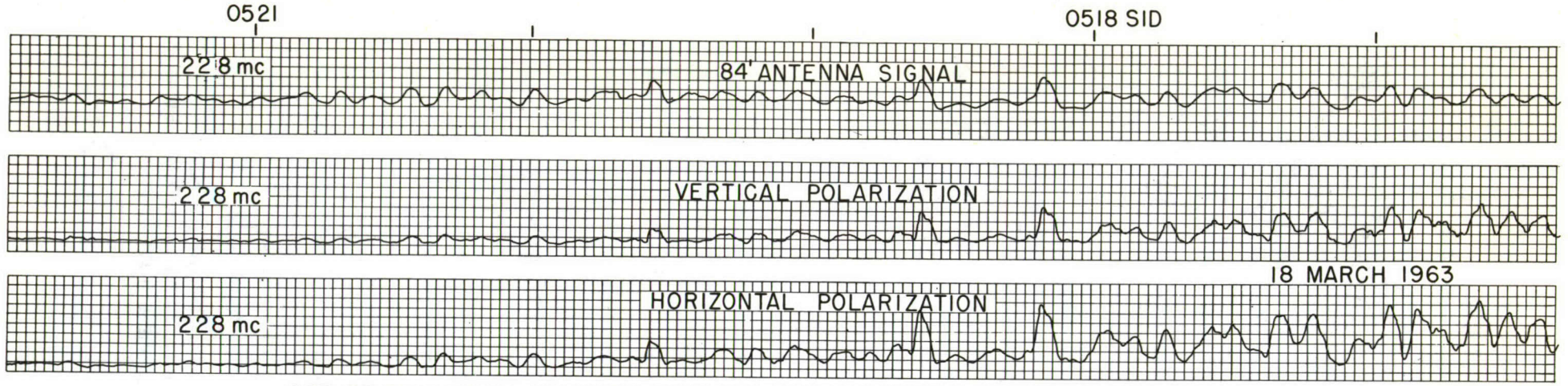
FIG. 1 LOCI OF INTERSECTION OF 400 KM IONOSPHERIC HEIGHT WITH THE RAY PATH TO CASSIOPEIA A AND CYGNUS A FOR THE SAGAMORE HILL RADIO OBSERVATORY. THE DOMAIN ACCESSIBLE TO SATELLITE STUDY IS ENCLOSED BY THE 0° ELEVATION RADIO HORIZON. ISOTOPES OF AURORAL ACTIVITY DIP TO THE WFST.





REV. 1238	EST.	LONG.	LAT.
MARCH 6, 1964	1928.00	66.92	20
1963-49C	1930.94	66.19	30
54 MC	1933.88	65.46	40
	1936.82	64.74	50

FIG. 2 PATH OF SATELLITE 1963 49-C GOING NORTH. FARADAY NULLS ARE CAUSED BY ROTATION OF THE PLANE OF POLARIZATION BY THE TOTAL ELECTRON CONTENT. ELECTRON DENSITY IRREGULARITIES CAUSE SCINTILLATIONS IN THE DISTURBED AURORAL ZONE



SET OF CYGNUS A: 84 FOOT ANTENNA AND 150 FOOT DUAL POLARIZED FEED:
FAST RUN RECORDING. NOTE CORRELATION OF INDIVIDUAL FLUCTUATIONS.

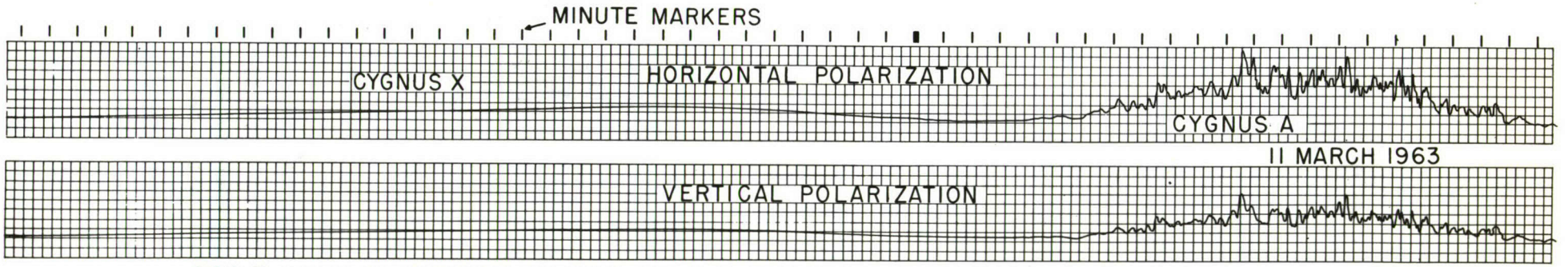


FIG. 3 SLOW RUN RECORDING OF CYGNUS A AND CYGNUS X AT 228 mc. CYGNUS X
DOES NOT SCINTILLATE BECAUSE OF ITS LARGE DIAMETER. 150 FOOT ANTENNA

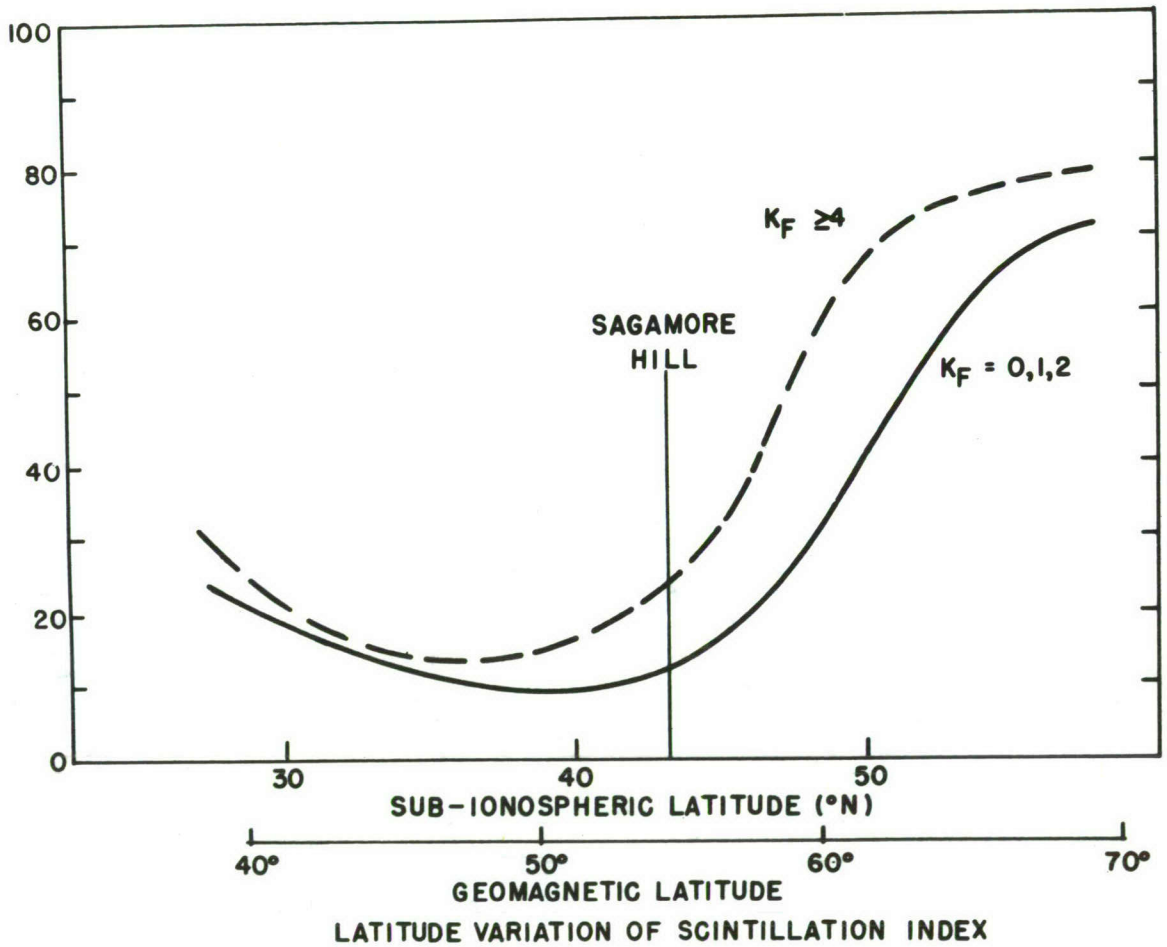


FIG. 4 SUB-IONOSPHERIC (400) KM LATITUDE DISTRIBUTION OF SCINTILLATION INDICES AT 54 Mc OF TRANSIT 4A DURING 1961 - 1962. NOTE SOUTHWARD MOTION AS MAGNETIC INDEX INCREASES DURING DISTURBED PERIOD. MAIN MAGNETIC STORM EFFECTS ARE AT HIGH LATITUDES

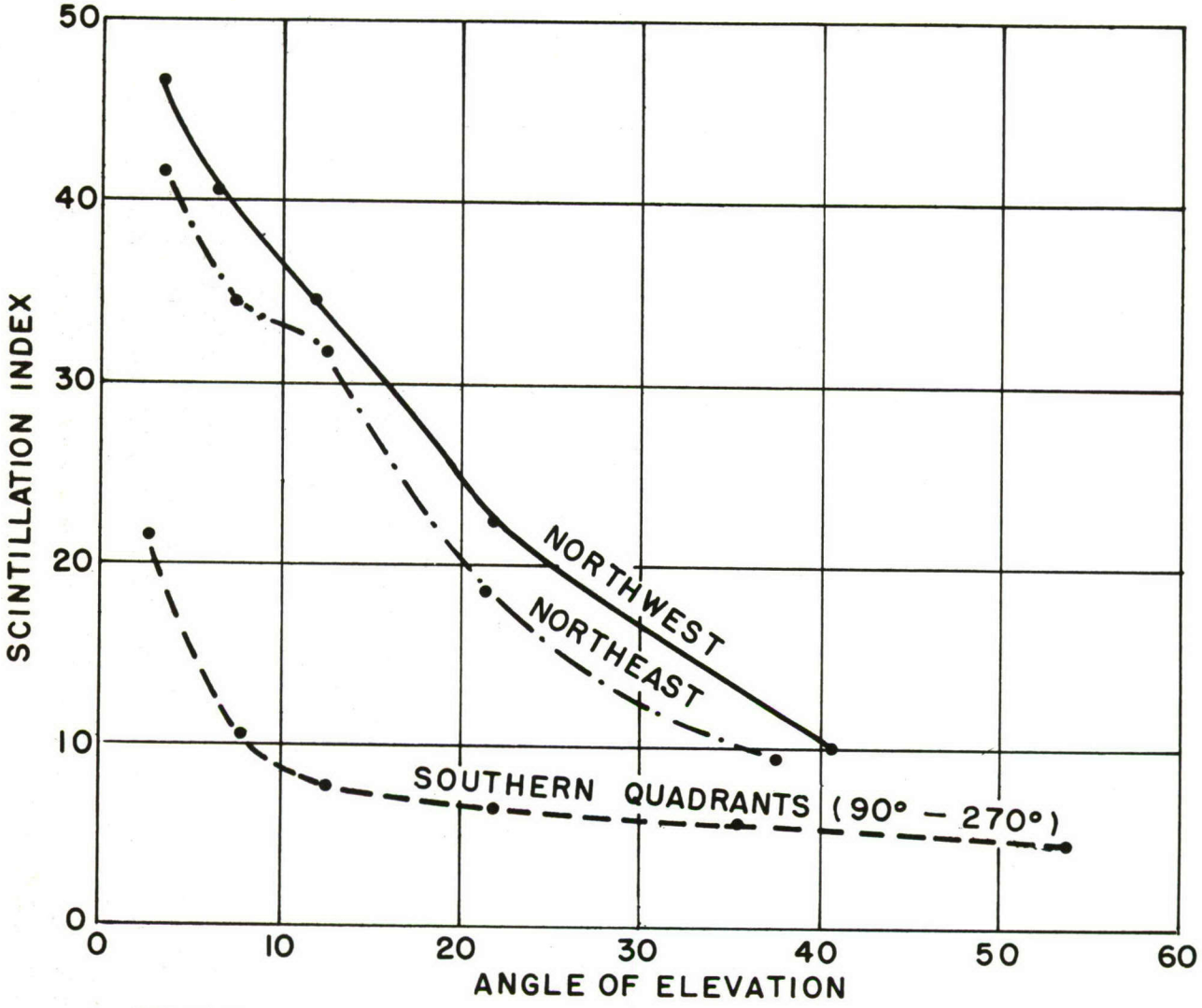


FIG. 5 SCINTILLATION INDEX AS A FUNCTION OF ANGLE OF ELEVATION SOUTH OF THE SAGAMORE HILL RADIO OBSERVATORY, NORTHEAST QUADRANT AND NORTHWEST QUADRANT. THE MAGNETIC DECLINATION OF THE OBSERVATORY IS 16° W.

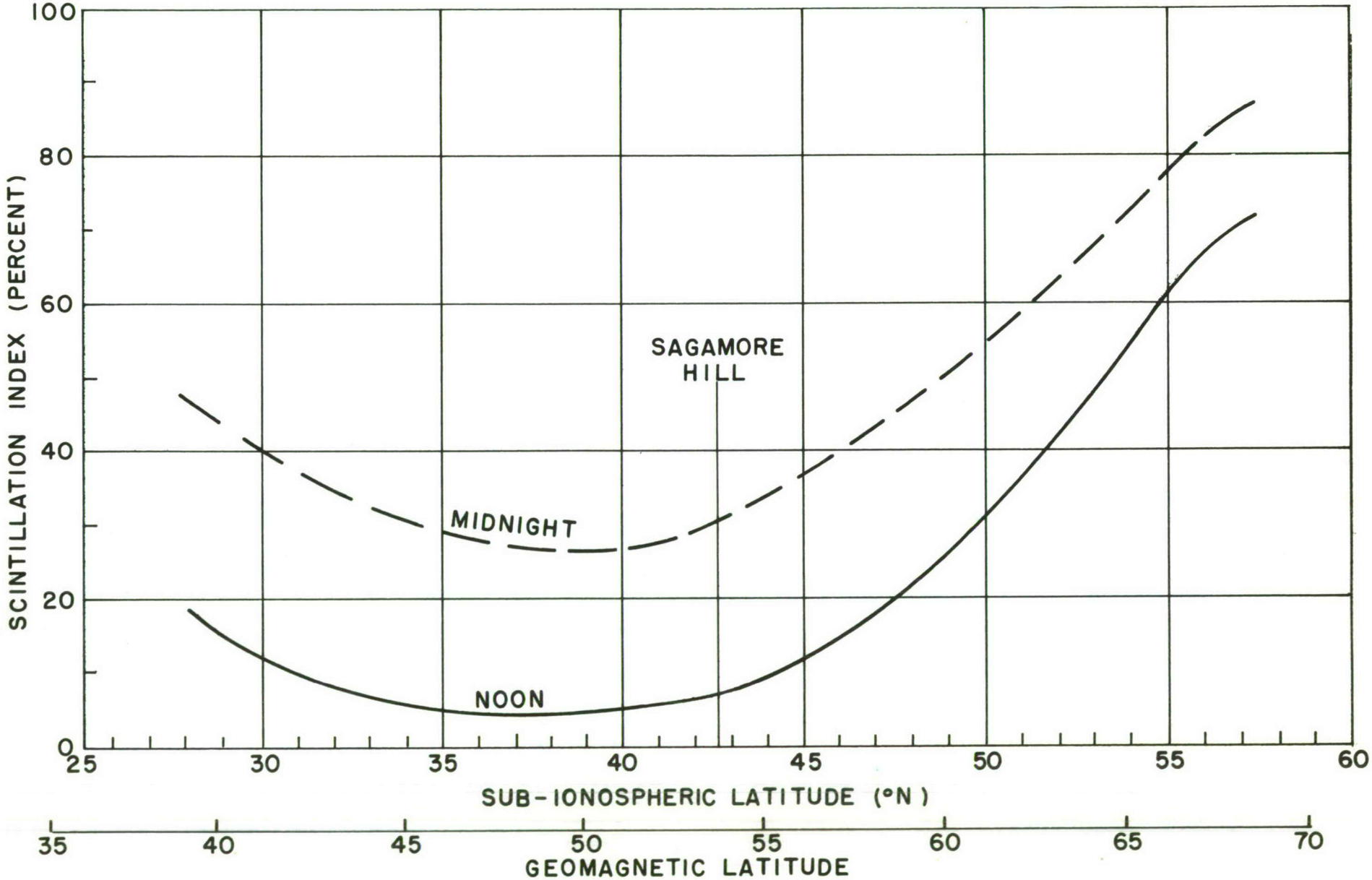


FIG. 6 DIURNAL VARIATION OF SCINTILLATION INDEX (K= 0, 1, 2)

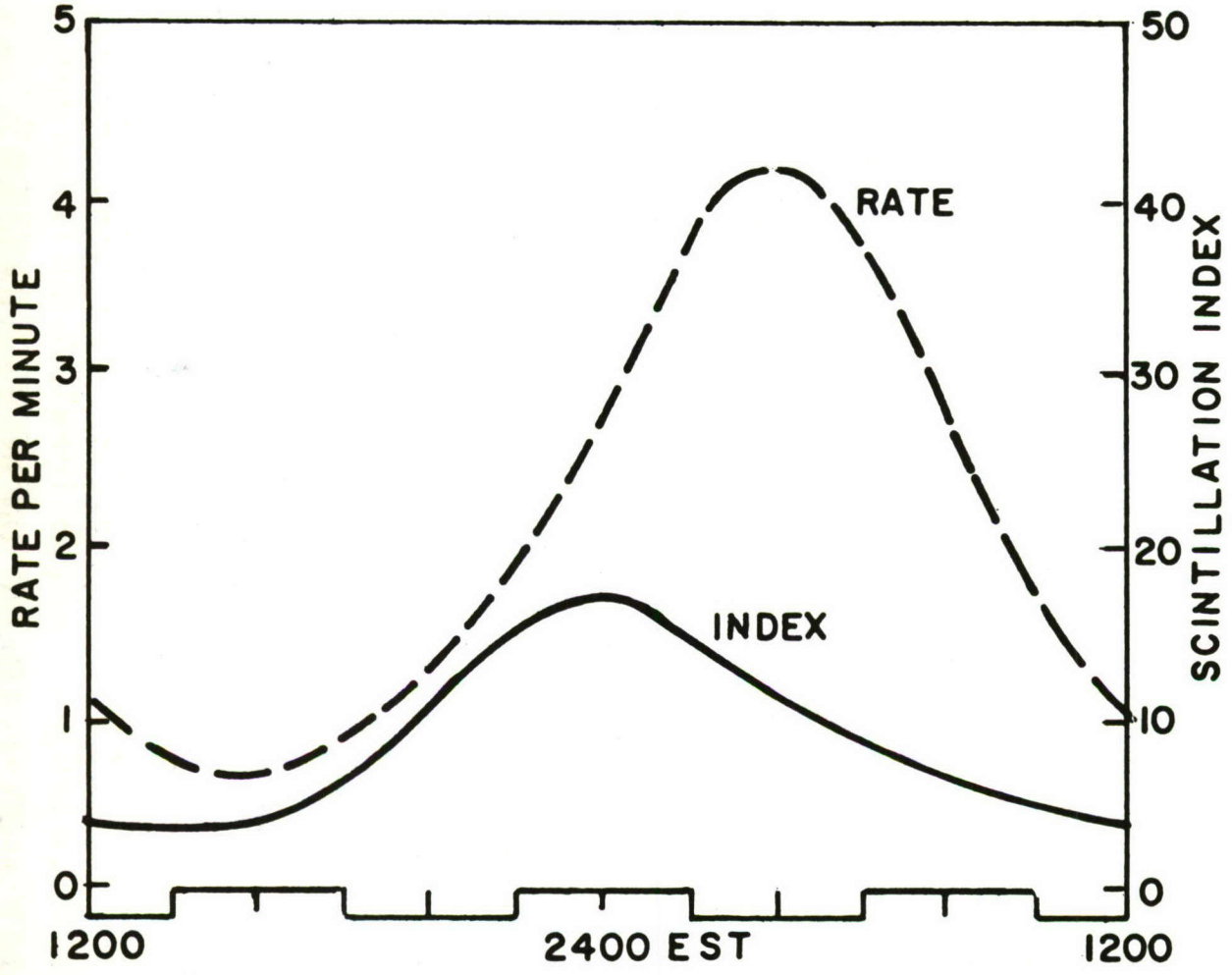


FIG. 7 DIURNAL BEHAVIOR OF 63 Mc SCINTILLATIONS FOR CASSIOPEIA A NEAR UPPER CULMINATION DURING 1961 AND 1962. MEAN ELEVATION ANGLE 60° FREDERICKSBURG MAGNETIC INDEX K_F = 0, 1

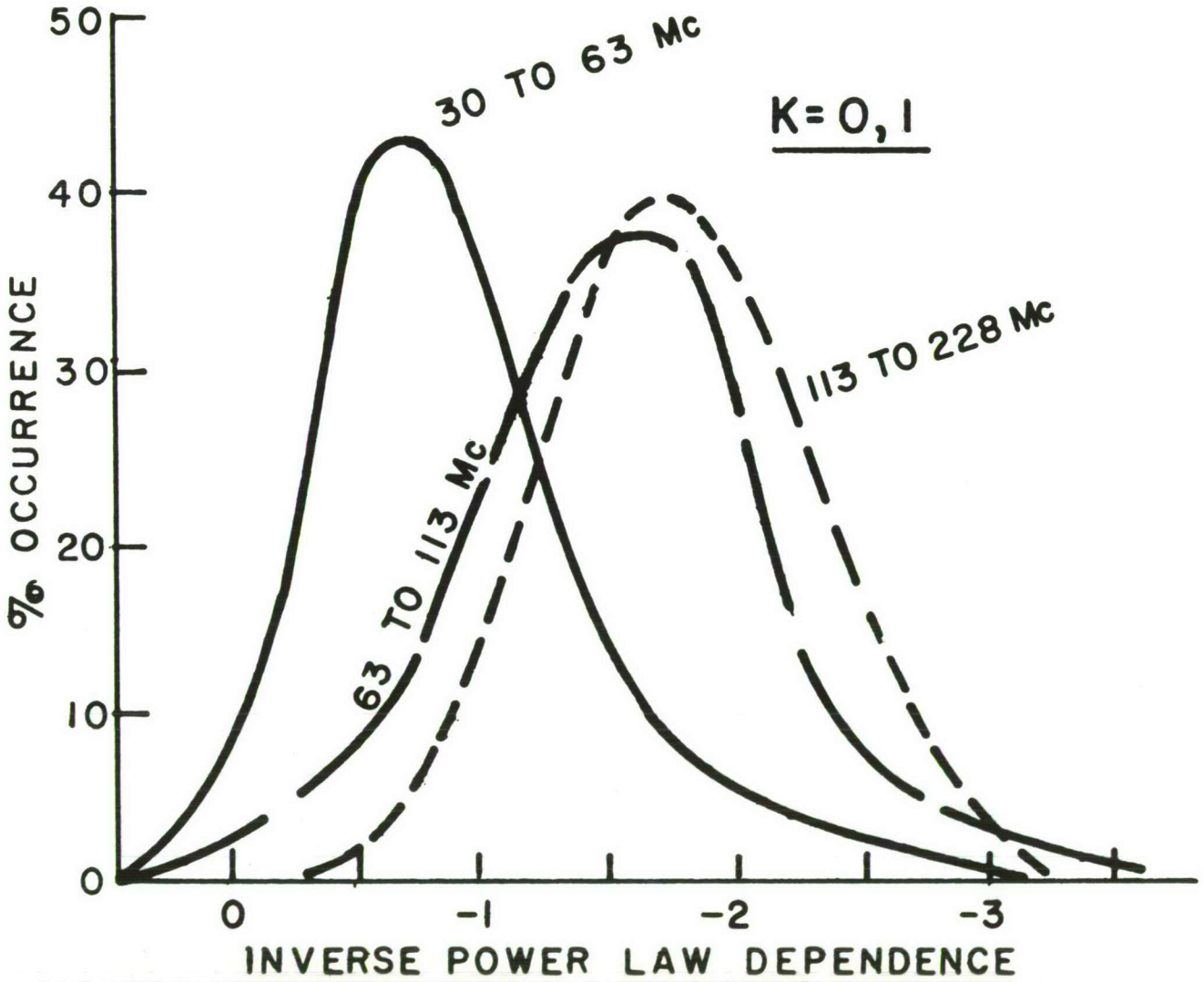


FIG. 8 FREQUENCY DEPENDENCE OF AMPLITUDE OF SCINTILLATIONS. CASSIOPEIA A DATA, JUNE 1963 TO JANUARY 1964 AT 12.5° ELEVATION

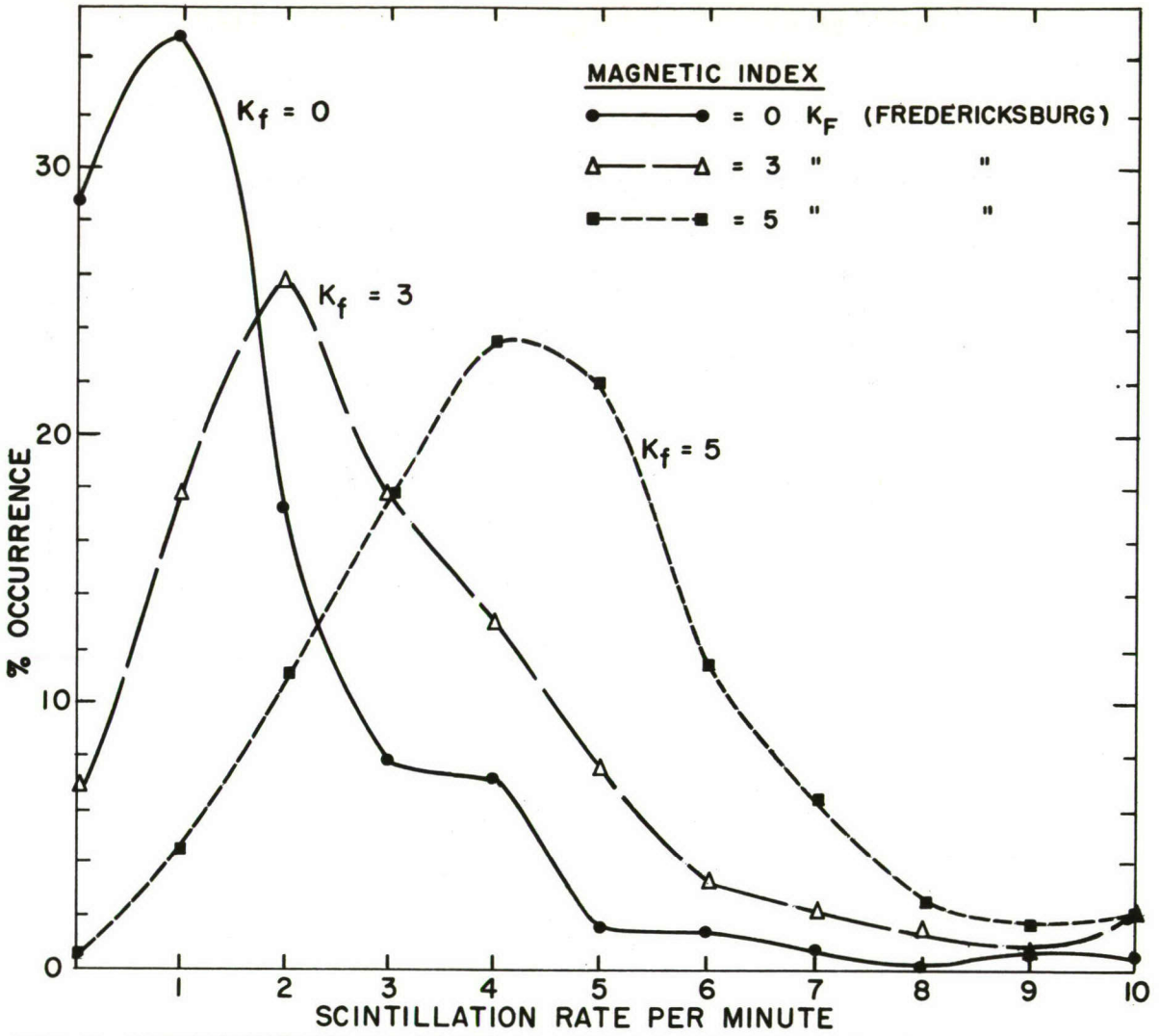


FIG. 9 DISTRIBUTION OF SCINTILLATION RATE PER MINUTE FOR VARIOUS STATES OF LOCAL MAGNETIC ACTIVITY. 1962 AND 1963 DATA FROM THE EQUATORIAL 84 FOOT ANTENNA, SAGAMORE HILL RADIO OBSERVATORY (2268 OBSERVATIONS)

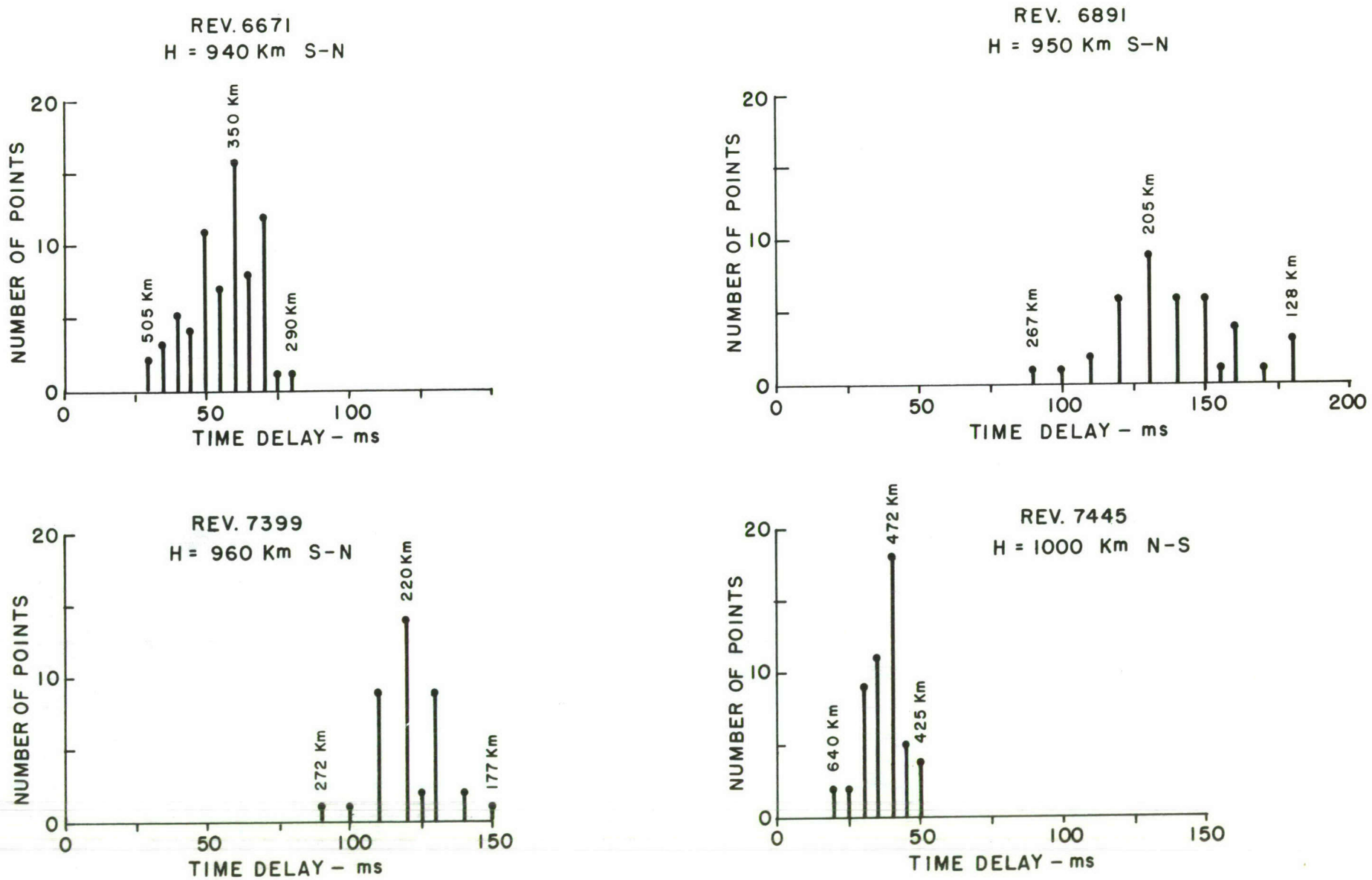


FIG. 10 HEIGHT DISTRIBUTIONS FOR SCINTILLATING REGIONS

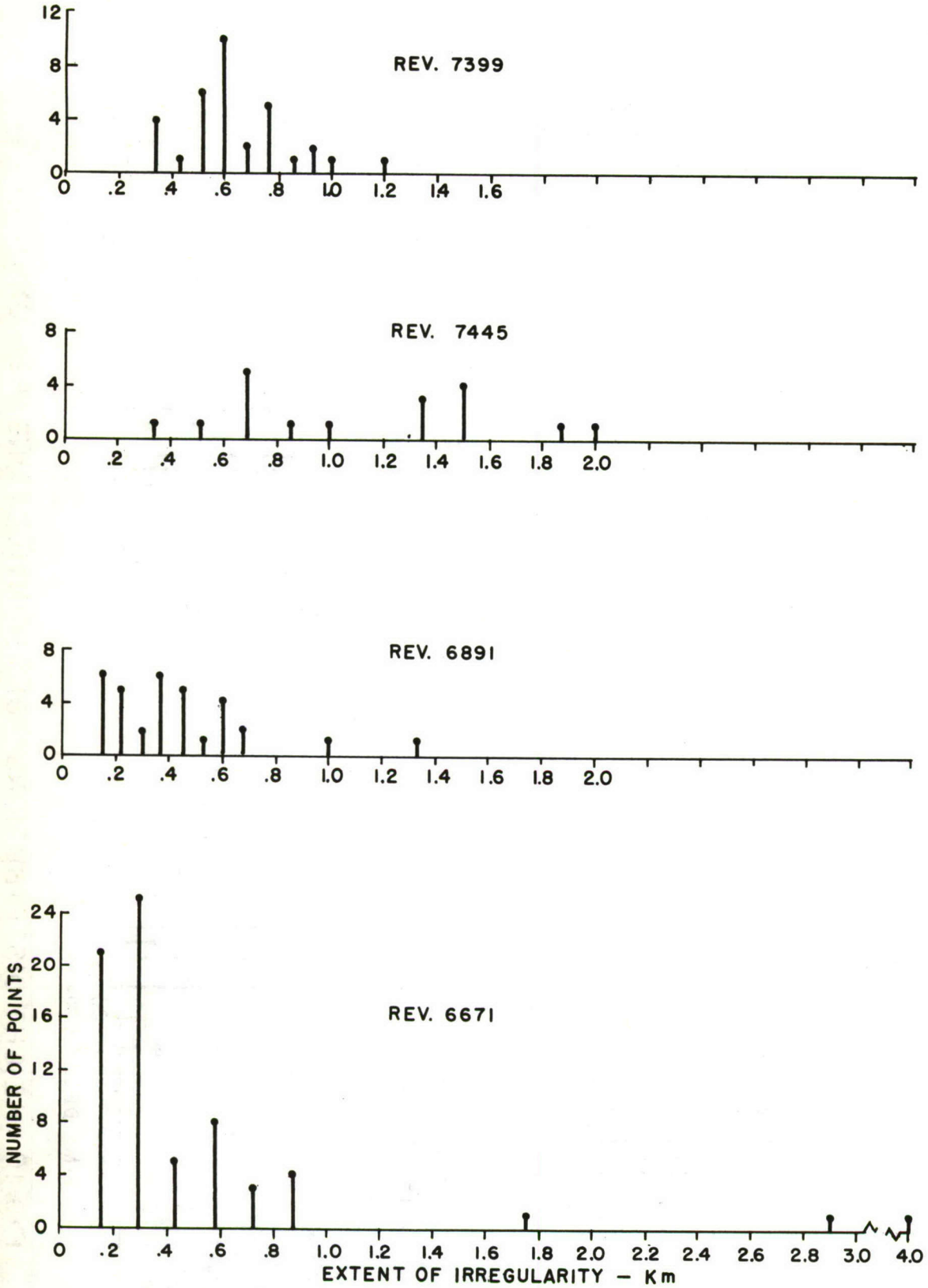


FIG. 11 DISTRIBUTION OF IRREGULARITY SIZES ALONG THE SATELLITE PATH.

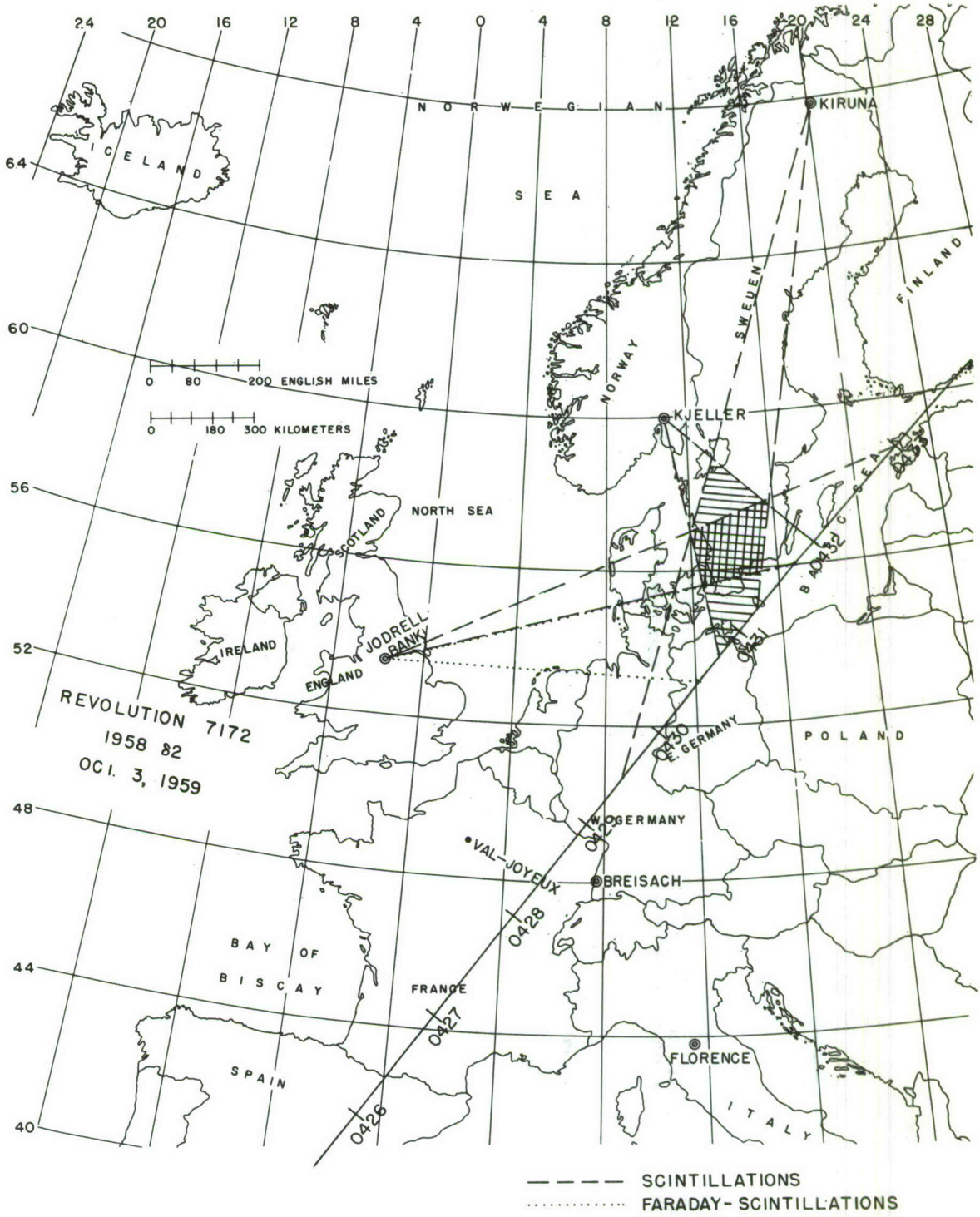


FIG. 12 SCINTILLATION REGION OBSERVED DURING 1958 Δ2
REVOLUTION 7172, 3 OCTOBER, 1959

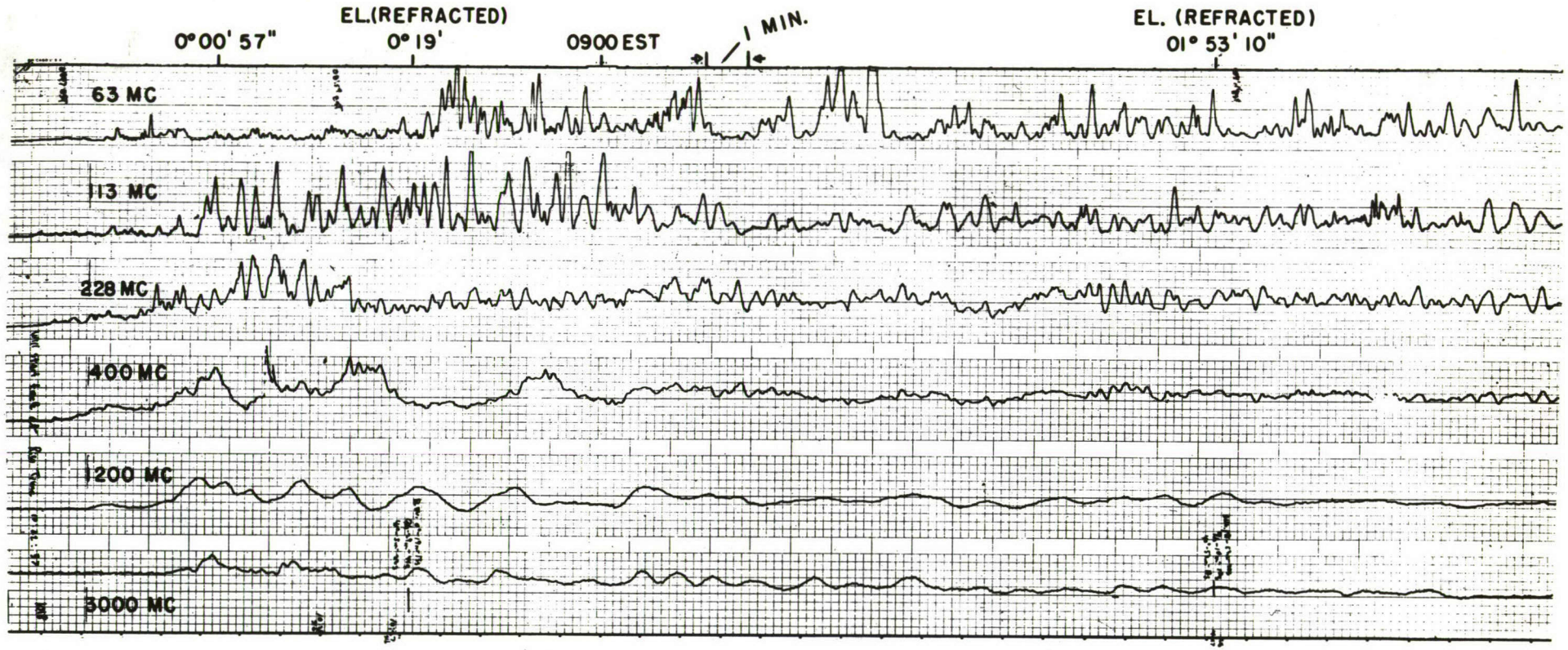


FIG. 13 OCTOBER 10, 1962 CYGNUS RISE K_F = 3+. IONOSPHERIC AND TROPOSPHERIC SCINTILLATIONS, 84' DISH, HAMILTON, MASS. MODERATE MAGNETIC ACTIVITY.

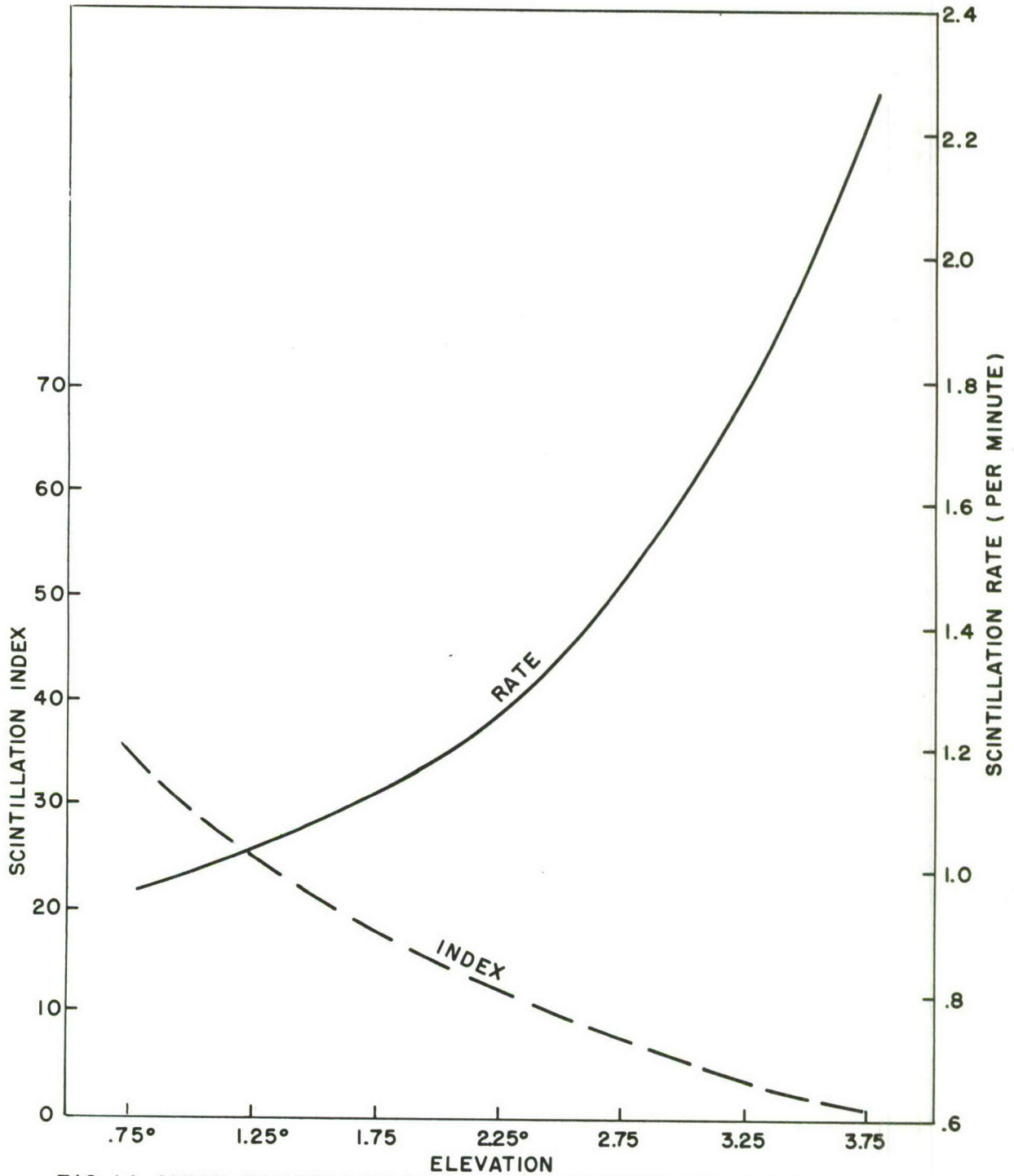


FIG. 14 MEAN SCINTILLATION INDEX AND RATE AT 1200 Mc AND 2980 Mc FOR CYGNUS RISE AND SET

ELECTRON RADIATION DAMAGE IN ZnSe AT LOW TEMPERATURE

by

B. A. Kulp and R. M. Detweiler
Aerospace Research Laboratories

Abstract

A threshold for radiation damage in ZnSe has been previously reported at 240 keV under electron bombardment. Fluorescence bands at 5460 and 5850 Å are produced. The damage causing these fluorescence bands anneals completely in a single stage at about 130°K. At 10°K somewhat different fluorescence bands are produced. Following heat treatment above 90°K the fluorescence spectrum produced by the 10°K bombardment anneals into that produced by the 80°K bombardment. This indicates that at least one and possibly more of the fluorescence centers observed following bombardment at 80°K is a "complex" center. That is, it is composed of a defect of Zn or Se in association with a foreign impurity. A second threshold for radiation damage in ZnSe has been observed while bombarding at 10°K. This threshold is at 190 keV and is believed to be that for displacement of the Zn atom from the lattice. The 240 keV threshold is therefore associated with displacement of the selenium atom. The displacement energies are 7.6 eV for the zinc and 8.2 eV for the selenium atom.

These values are compared to 7.3 eV and 8.7 eV for the cadmium atom and sulfur atom previously measured in CdS.

Introduction

In the study of solids one of the most interesting and important topics is the role of native defects in determining the properties of the material. Specifically, lattice vacancies and interstitials of the host atom species often determine the useful properties of the material or conversely may seriously limit the utility of a particular material. There are several methods of producing vacancies and interstitials in materials. A common method employed is to heat the material to a temperature near its melting point, then quench it to a low temperature, thus "freezing in" the high temperature equilibrium concentration of defects. Such treatment has the major disadvantage that all the defects as well as the other impurities are able to diffuse rapidly through the material, and hence the configuration of the defects is not known. That is, the defects may form clusters or complexes with each other or with other chemical impurities. Another method of producing defects applicable to the alkali halides is by irradiation with ultraviolet light or x-rays. This method is quite convenient if one is studying a material where such defect producing mechanisms work, but is not generally applicable to all materials. A third method of producing vacancies and interstitials is by electron bombardment. If the electrons are of an energy slightly above the threshold for displacement of an atom from the lattice, then only one vacancy and one interstitial will be produced in a single

collision. If the temperature is low enough (generally below about 25°K for semiconductors, or below 15°K for metals) there will be no diffusion of the defects and it is possible to produce simple isolated vacancies and simple isolated interstitials. Furthermore, if one is studying a material such as CdS where the atomic weight of the two species differs somewhat, it should be possible to observe two thresholds for displacement. Thus, by bombarding between the two, one can produce only defects of the lighter species and by bombarding at energies above the higher threshold one should produce defects of both species.

A further advantage of electron bombardment for producing the defects is that one can control the number of the defects produced by controlling the number of electrons striking the sample.

In particular, consider the II-VI compound semiconductors composed of elements which appear in Columns II and VI of the Periodic Table. They consist of the oxides, sulfides, selenides and tellurides of zinc, cadmium and mercury. At the Aerospace Research Laboratories, OAR, effort is directed primarily in the field of CdS, ZnS, CdSe and ZnSe. These materials are useful fluorescent materials and possess useful photoconductive and photovoltaic properties. More recently, piezoelectric properties of CdS have been shown to be applicable to ultrasonic amplification. Large single crystals of these materials (of the order of cm^3) are grown by vapor phase deposition and often have very brilliant fluorescences or unusual electrical properties or both which can only be explained by native defects,

since chemical impurities can be reduced below the limit of detectability by ordinary emission spectroscopy. The lack of stoichimetry in a partifular crystal is not sufficiently great to detect chemically. Thus, typically we might expect that in a crystal which has too much cadmium, the excess Cd will be of the order of 0.01%, not of the order of 1 to 10 percent which is often observed in certain ceramic materials. The program in electron irradiation was started at ARL in order to study specifically the stoichimetry problem in II-VI compound semiconductors. The first experimental observation of radiation damage in CdS was made by Reynolds[1] of ARL when he produced a green fluorescence observable at 77°K under ultraviolet light stimulation by bombardment with 1 meV protons. Subsequently, the threshold for the production by electron bombardment of the defect responsible for this fluorescence was measured by Kulp and Kelly[2] to be 115 keV. This threshold has been interpreted as the electron energy necessary for the displacement of a sulfur atom, also, the defect responsible for the fluorescence has been identified as a sulfur atom in an interstitial (non-lattice point) position. A second threshold for "damage" has been measured in CdS at 290 keV[3], which is the threshold for the displacement of the cadmium atom from the lattice. The present material under investigation is ZnSe. The current results of this work are presented here.

Experimental Methods

The crystals used in these experiments were grown by the vapor phase deposition method of Reynolds[4] in a furnace shown schematically in Fig. 1. Since there is a temperature gradient in the furnace, the charge is placed in the center of the furnace at the hottest point. For large crystals, seed plates are used and are placed at a position in the furnace where there is optimum growth of large perfect crystals. For plate and whisker type crystals, no seed plates are used and the crystals grow on the ends of the charge or on the sharp edges of slots cut into the inner quartz tube. The maximum temperature in the furnace for the ZnSe crystals used in these experiments was in the range 1250 to 1350°C. The crystals were generally about 1 mm^3, and several were placed on the target at one time. The crystals had the zinc blende structure, but were somewhat stack-faulted.

The accelerator used for the electron bombardment is a Van de Graaff type accelerator with an energy range of 200 keV to 1 meV. Figure 2 shows the overall accelerator installation with the cryostat for carrying out bombardment at liquid helium temperature. The accelerator is in the background. The cryostat is in the right foreground. The Perkin Elmer glass prism spectrometer, in the center foreground, with a type 7102 photomultiplier was used for the quantitative intensity measurements in the threshold and the annealing experiments. The Bausch and Lomb grating spectrometer in the left foreground with Kodak Type 103-aF plates was

used for the quantitative measurements of the wavelengths of the various fluorescence bands observed.

Figure 3 shows the details of the helium cryostat. The helium container is constructed entirely of welded stainless steel. The helium valve can be closed and the liquid evaporated out of the cold finger for the heat treatment experiments. The temperature is measured with a platinum resistance thermometer (Rosemont type 118L.). An a.c. resistance bridge was used to measure its resistance.

Results

a. The threshold for displacement.

A quantity of fundamental importance in the use of the electron beam to produce defects in a material is the threshold for the displacement of an atom from the lattice. In the experiments on II-VI compounds, displacement is observed by a change in the fluorescence spectrum of the material. Two distinct types of ZnSe crystals have been found under electron bombardment. Figure 4 shows the fluorescence of "type A" ZnSe crystals at 80°K before and following bombardment at 80°K by 500 keV electrons.[5]

The peaks at 5460 and 5850Å are produced by the bombardment. The intensity of the peak at 5460Å increases linearly with the number of electrons striking the crystal. Figure 5 shows the rate of increase of the fluorescence as a function of electron energy for bombardment at 80°K. This rate of increase is proportional to the cross-section for displacement* so that the intercept with the energy axis at 240 keV is the threshold for displacement of an atom from the lattice.

* $I \propto N_c$

and $dN_c/dt = No\ n_e\ v\sigma$,

where I = intensity of fluorescence, N_c = number of centers, n_e = number of electrons per second (a constant), N_o = number of atoms in crystals (a constant), v = electron velocity (energy) (a constant) and σ = displacement cross section. Therefore, $\sigma \propto dI/dt$.

An electron of this energy transfers a maximum energy of 8.2 eV to a selenium atom and 10 eV to a zinc atom. The former is more in line with the displacement energies observed in CdS.

At 10°K a different fluorescence spectrum is produced by electron bombardment. Figure 6 shows the fluorescence spectrum at 10°K following electron bombardment at an energy of 500 keV and a temperature of 10°K, as well as the spectrum at 10°K of the same crystal when bombarded at 80°K. Figure 7 shows the rate of increase of intensity of the damage at 10°K as a function of electron beam energy. The threshold in this case is at 190 keV, significantly lower than at 80°K. The 190 keV electron energy would be sufficient to transfer 6.3 eV to a selenium atom and 7.6 eV to a zinc atom.

"Type B" ZnSe crystals show no change in its fluorescence spectrum in the wavelength region 0.5 to 2.5μ during electron bombardment at 80°K. However, when bombarded at 10°K, these crystals do show radiation damage with a threshold of 190 keV.

B. Annealing of the Radiation Damage

In addition to the threshold for radiation damage, the annealing characteristics are of prime importance. Figure 8 shows the isochronal annealing of a crystal bombarded at an energy of 500 keV and a temperature of 10°K. This type annealing consists of heating the crystal to the desired temperature, holding at that temperature for a fixed period of time (10 minutes in these experiments), then cooling to the reference temperature. Thus, all measurements are made at the same temperature and the effect of temperature alone

on the spectrum is not considered. In Figure 8, three stages of annealing are very apparent, one at about 60°K, another at 90°K, and a final one at 130°K. The annealing is complete and following a heat treatment at 130°K or above it is necessary to bombard above the threshold in order to produce the fluorescence again. There is some additional annealing between 60° and 90°K but only very slowly. When crystals are bombarded at 80°K, however, the annealing consists essentially of a single stage at 130°K. In type B crystals, only the stages at 60° and 90°K are observed. The annealing is complete following the 90°K stage. The annealing characteristic of crystals bombarded below 240 keV, but above 190 keV consists of a single stage at 90°K.

The fluorescence spectrum following each of the annealing stages is important because it tells if the damage remaining is of the same type as that before the annealing stage. Figure 9 shows the spectra of the crystal initially and following each of the first two annealing stages after bombardment at 10°K. The spectrum following the 60°K stage is not only reduced in intensity but is somewhat different in shape than that before the anneal. Finally the spectrum following the 90°K stage is very different from that at the previous stages.

The effect of temperature alone on the spectrum is demonstrated in Fig. 10, which shows the spectra taken at both 10°K and 80°K, of a crystal following bombardment at 10°K and heat treated at 95°K, as well as the spectrum at 80°K following bombardment at 80°K. Note

that spectra 2 and 3 are very similar except for the long wavelength peak, and that they are essentially the same as spectra B in Fig. 4. The long wavelength peak does not show up in curve 2 or B probably because of different bombardment times which result in different relative intensities of the various fluorescence bands.

C. Kinetics of the Annealing Process.

Meechan and Brinkman[6] have derived a set of relations which enable a determination of the order of the reaction and the activation energy of the process of annealing. Their method utilizes the isochronal and a single isothermal annealing characteristic on the same or very similar samples. Figure 11 shows a plot of ln ΔT, while it is the "equivalent" time of annealing at a given temperature as a function of 1/T for the three stages of annealing. The slope of these curves gives the activation energies for the annealing processes.

Another characteristic of the annealing processes useful in their identification is the order of the reaction. This may be obtained from the isothermal annealing curves by the method for Meechan and Brinkman. Figure 12 shows a log-log plot of the intensity vs the time for three isothermal anneals. The slope of the line plus one is the order of the process. The anneals at 60 and 90°K are clearly second order reactions while that at 130°K is of considerably higher order. Table 1 summarizes the activation energies and the order of the process for the three annealing stages described here.

Discussion of Results

Threshold for Displacement

Consider first the thresholds for the production of radiation damage in ZnSe. A 240 keV electron transfers a maximum of 8.2 eV to a selenium atom and 10 eV to a zinc atom while a 190 keV electron transfers 6.3 eV to a selenium atom and 7.6 eV to a zinc atom. Comparison to the CdS case when the displacement energies were found to be 7.3 eV and 8.7 eV for the cadmium and sulfur atoms respectively, leads to the conclusion that the 240 keV threshold most probably is that for the displacement of the selenium atom, while the 190 keV threshold represents that for the zinc atom. The displacement energies are then 8.2 eV for the selenium atom and 7.6 eV for the zinc atom. The difference is considerably greater than the experimental error which is about 5 percent (relative). It is, however, consistent with the results obtained from CdS where the cation was more easily displaced than the anion.

Annealing

From the isothermal annealing characteristics, the following rather dramatic observations can be made. The fluorescence due to both the selenium and zinc vacancies and interstitials are at nearly the same wavelengths however th ey can be separated by their annealing characteristics. From Fig. 6 it is seen that the fluorescence produced by the selenium vacancies and interstitials is observable at 10°K following bombardment by electrons with energies greater than 240 keV. At about 60°K one of these defects diffuses through the

lattice with an activation energy of 0.35 eV and in type A crystals finds another impurity, or impurities, to form the complex centers responsible for the fluorescence spectra following bombardment at 80°K above the threshold energy of 240 keV. There is no evidence of the formation of divancies or complexes involving zinc defects since there are crystals where such complexes are not observed. It is not possible at the present time to determine just how many complexes are formed because the temperature dependence of the fluorescence bands have not been fully investigated but there is at least one complex, or possibly three, if one considers all of the additional fluorescence bands observed in Figs. 6 and 10.

The annealing stage at 130°K is also very significant. It may represent either the decomposition of the complex center or it may represent the onset of diffusion of the other fundamental defect of selenium, the selenium vacancy, which finds the selenium interstitial originally produced and recombines with it. The latter explanation seems more probable, since in various as-grown crystals, the fluorescence bands shown here have been observed which indicate that the complex centers themselves may be stable to rather high temperature (a few hundred degrees C). The activation energy for the latter process is found to be 0.4 eV.

Finally, the annealing stage at 90°K represents the annihilation of the zinc defects produced by bombardment at 10°K by electrons of energy greater than 190 keV. This process takes place with an activation energy of 0.15 eV. As yet it is not possible to decide whether

this represents the diffusion energy of the vacancy or the interstitial. In the case of metals, however, it is now generally believed that the interstitials move through the lattice with a lower activation energy than the vacancies.

A factor in semiconductors not present in metals which may contribute to the ease or diffulty of diffusion of the interstitial is the charge on the particle. In the case of the cation, it probably is not so important since the cation will tend to lose electrons thereby decreasing its effective radius slightly. In the case of the anion however, an increase in the negative change, of selenium for example, enlarges the ion and may make it more diffult for it to squeeze from one interstitial position to another. Such factors are yet to be investigated.

Conclusion

It has been demonstrated that by using electron radiation damage it is possible to identify fluorescence bands in ZnSe (and other compounds) with defects of the selenium atom and of the zinc atom as well as with complexes involving a defect of selenium and a foreign impurity. In addition, activation energies for diffusion of certain of these defects have been measured. While it is not possible in some of these instances to identify the defect exactly, the ability to produce a given defect at will and to control the number density of these defects is indeed a major step in such identification. It is believed that the electron beam has been established as a valuable new tool for use in the study of defects in II-VI compound semicon-

ductors and that the use of it in association with fluorescence studies, photoconductive properties and transport properties of the charge carriers will greatly add to the understanding of these materials and hasten and improve their future utility in the field of light production, photovoltaic, photoconductive and photoresistive devices.

List of Figures

Fig. 1 Schematic of crystal growing furnace.

Fig. 2 Van de Graaff Accelerator installation with Cryostat and Spectrometers.

Fig. 3 Cross-section of helium cryostat.

Fig. 4 Fluorescence spectra of type A ZnSe, A before and B after bombardment at 80°K with 500 keV electrons.

Fig. 5 Rate of increase of fluorescence as a function of electron energy for bombardment at 80°K. Note expansion of intensity scale by factors of 10 and 100.

Fig. 6 Spectra after electron bombardment at 500 keV 1. Bombarded and observed at 10°K. 2. Bombarded at 80°K and observed at 10°K.

Fig. 7 Rate of increase in fluorescence as a function of electron energy for bombardment at 10°K.

Fig. 8 Isochronal anneal of electron radiation damage in type A crystals, 10 min. at each temperature.

Fig. 9 Spectra of type A crystal following irradiation at 10°K 1. Before heat treatment (intensity reduced by factor of 2.5), 2. Following heat treatment at 65°K for 10 min., 3. Following heat treatment at 95°K for 10 min.

Fig. 10 Fluorescence at 10°K (1) and 80°K (2) of type A crystal bombarded at 10°K and annealed at 95°K. 3 Fluorescence at 80°K of crystal bombarded at 80°K.

Fig. 11 Meechan - Brinkman plot to determine activation energies.

Fig. 12 Plot to determine the order of the reactions for annealing.

References

1. D.C. Reynolds and L. C. Greene, Bull. Am. Phys. Soc. Ser. II, 3, 108 (1958).
2. B. A. Kulp and R. H. Kelley, J. Appl. Phys. 31, 1057 (1960).
3. B. A. Kulp, Phys. Rev., 125, 1865 (1962).
4. D. C. Reynolds and S. J. Czyzak, Phys. Rev., 19, 1957 (1950).
5. B. A. Kulp and R. M. Detweiler, Phys. Rev., 129, 2422 (1963).
6. C. J. Meechan and J. A. Brinkman, Phys. Rev., 103, 1193 (1956).

Table 1

Temperature of Annealing Stage	Activation Energy	Order of Reaction
60°K	0.35 eV	2
90°K	0.15 eV	2
130°K	0.4 eV	4-5

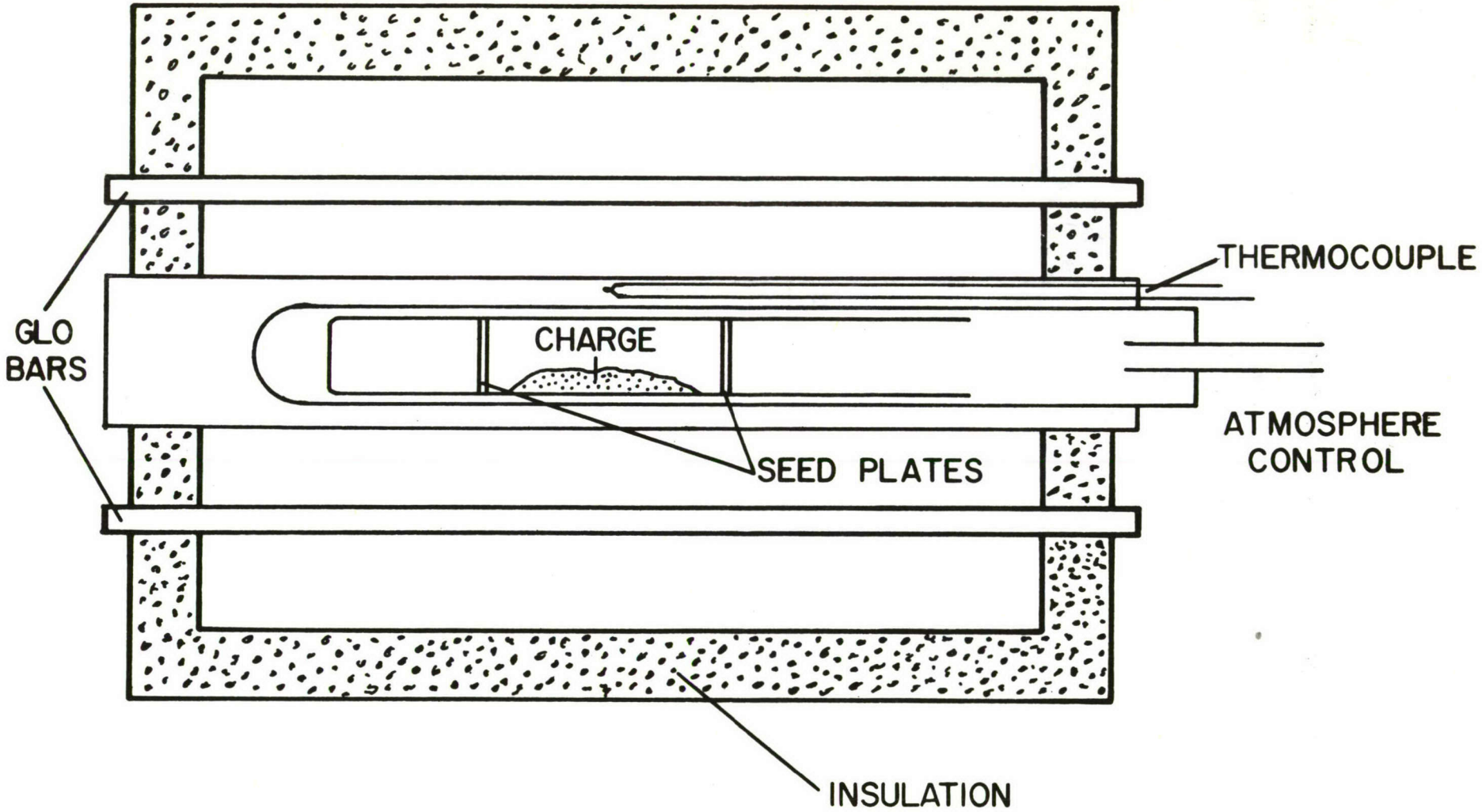

GLO
BARS
THERMOCOUPLE
CHARGE
SEED PLATES
ATMOSPHERE
CONTROL
INSULATION

FIGURE 1

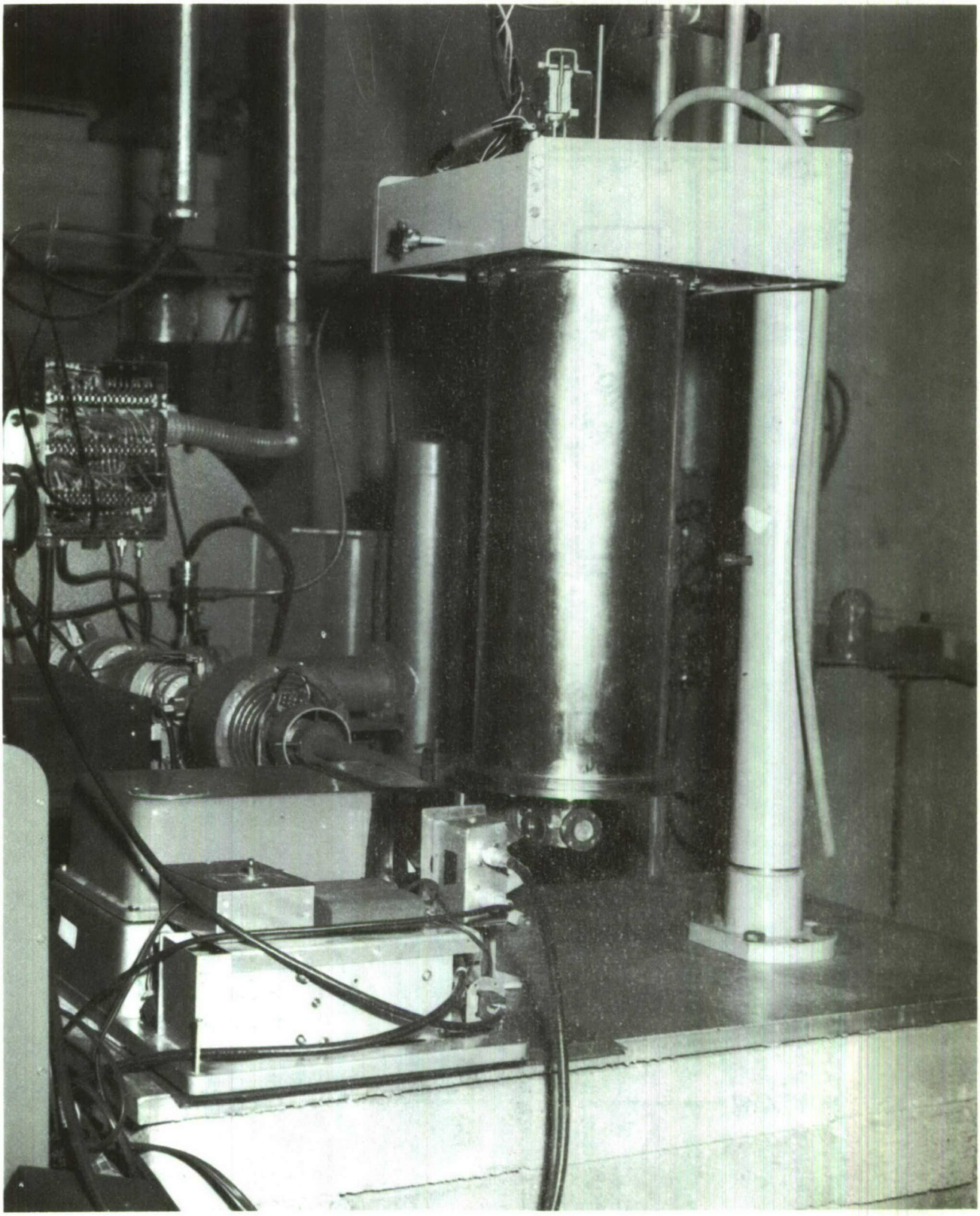


FIGURE 2

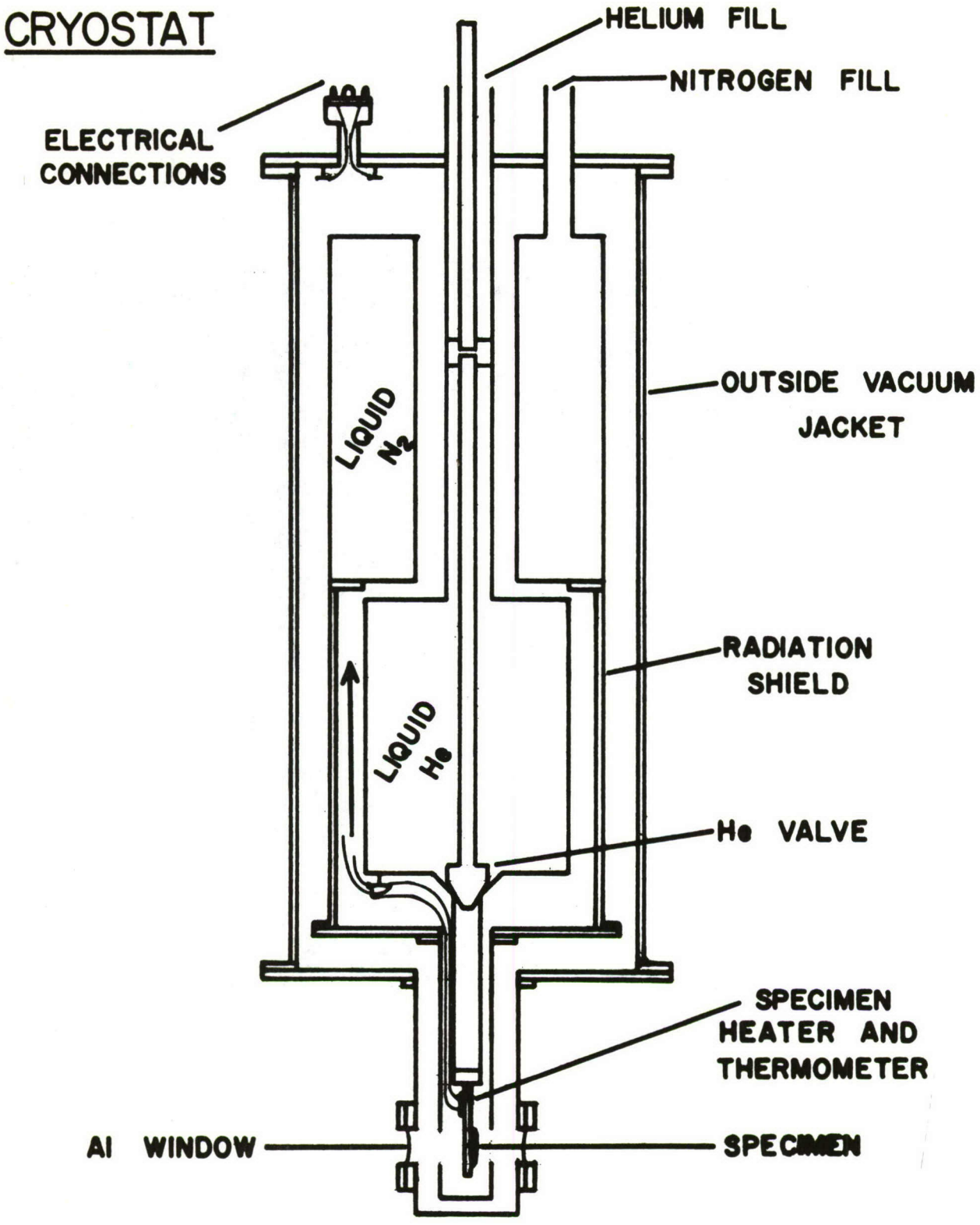

CRYOSTAT
HELIUM FILL
NITROGEN FILL
ELECTRICAL CONNECTIONS
OUTSIDE VACUUM JACKET
LIQUID N_2
RADIATION SHIELD
LIQUID He
He VALVE
SPECIMEN HEATER AND THERMOMETER
Al WINDOW
SPECIMEN

FIGURE 3

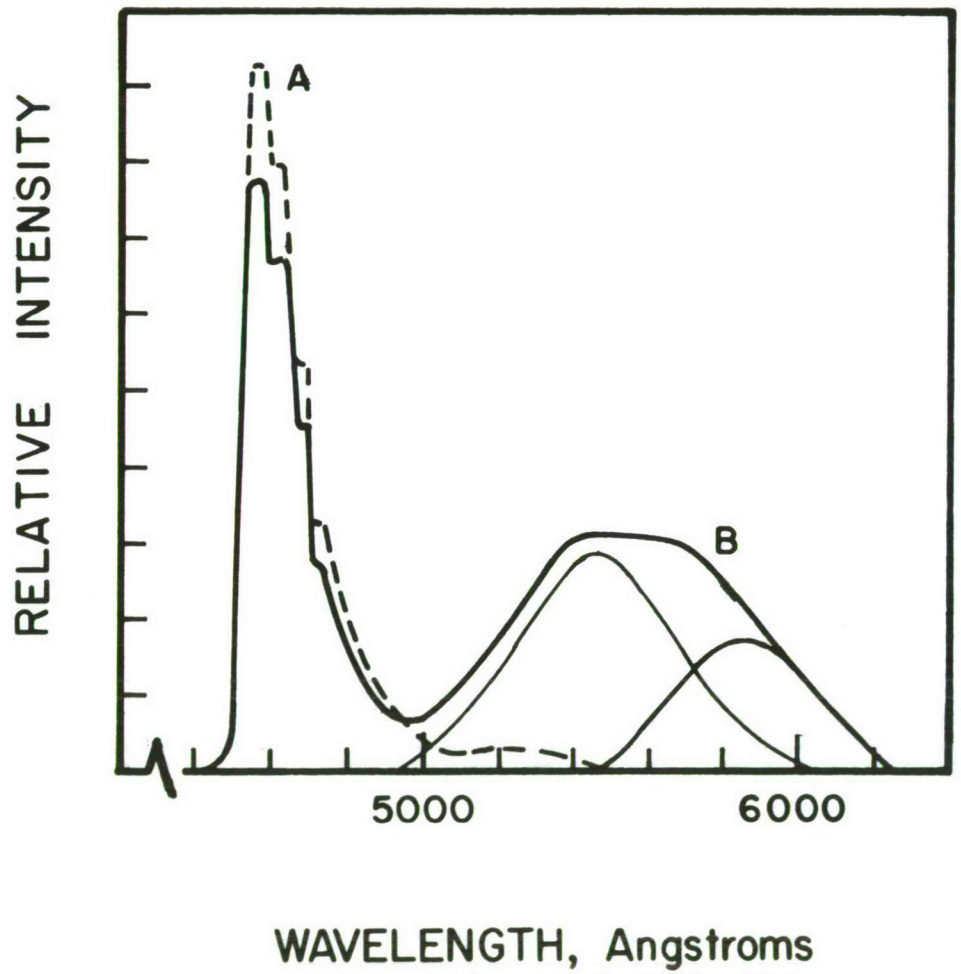


FIGURE 4

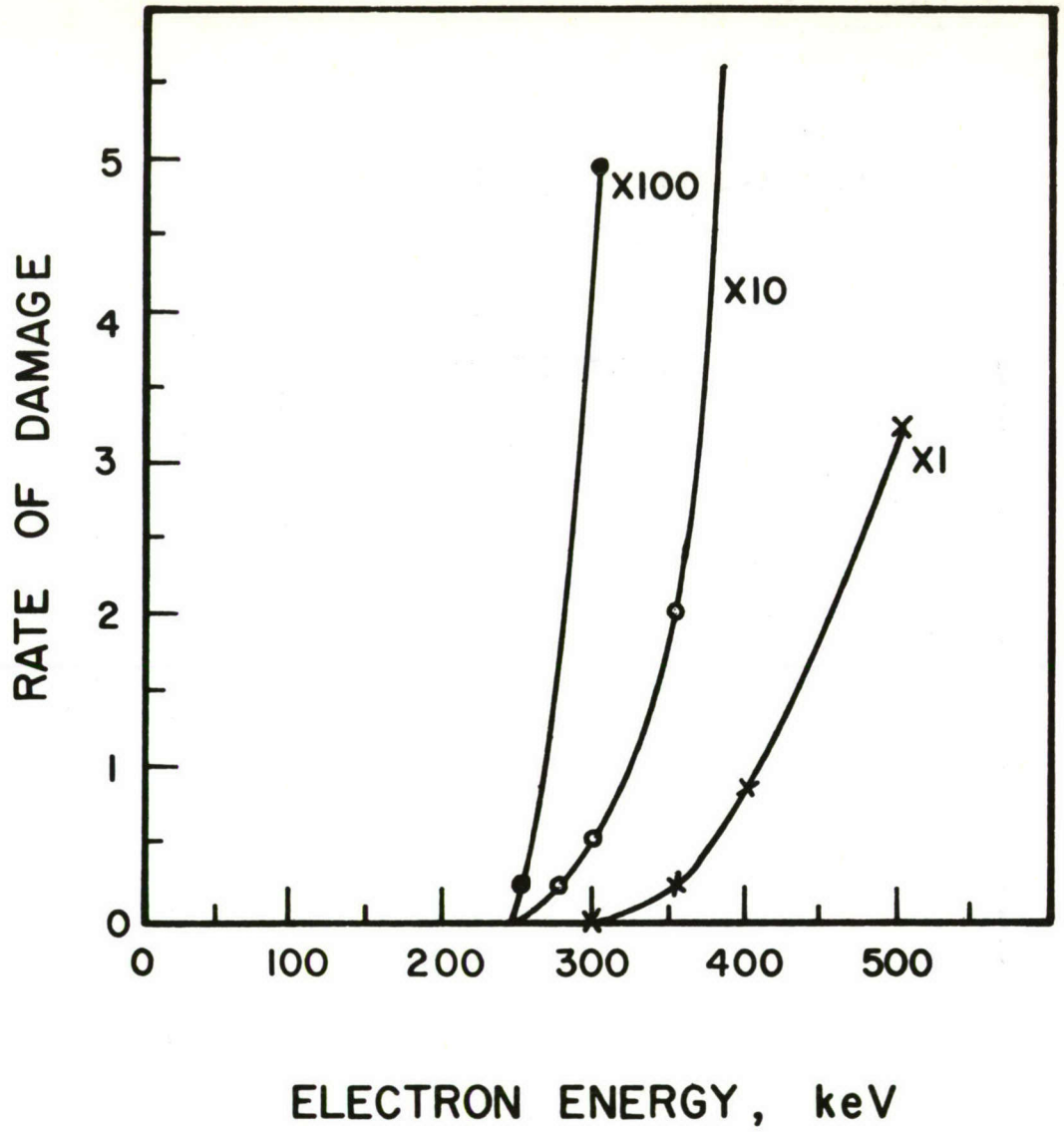

RATE OF DAMAGE
0
1
2
3
4
5
X100
X10
X1
0
100
200
300
400
500
ELECTRON ENERGY, keV

FIGURE 5

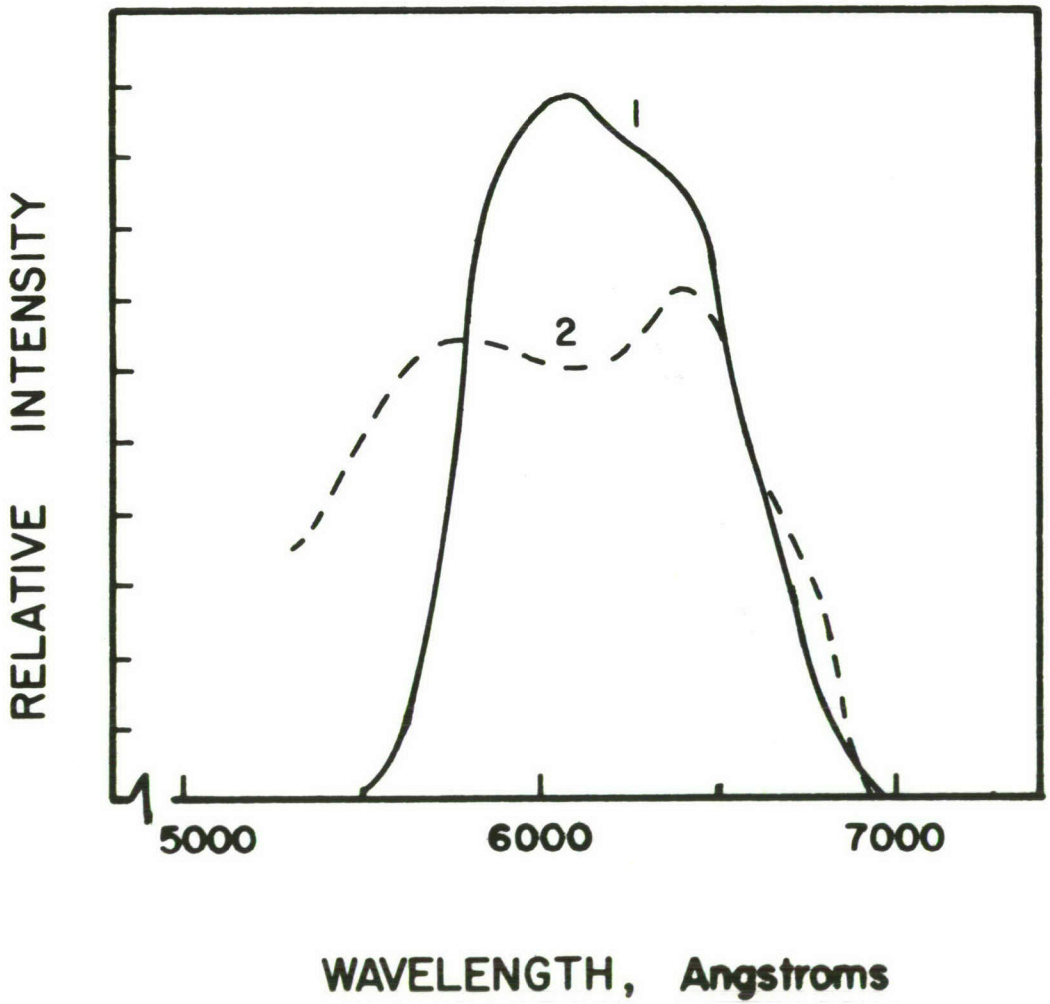

RELATIVE INTENSITY
1
2
5000
6000
7000
WAVELENGTH, Angstroms

FIGURE 6

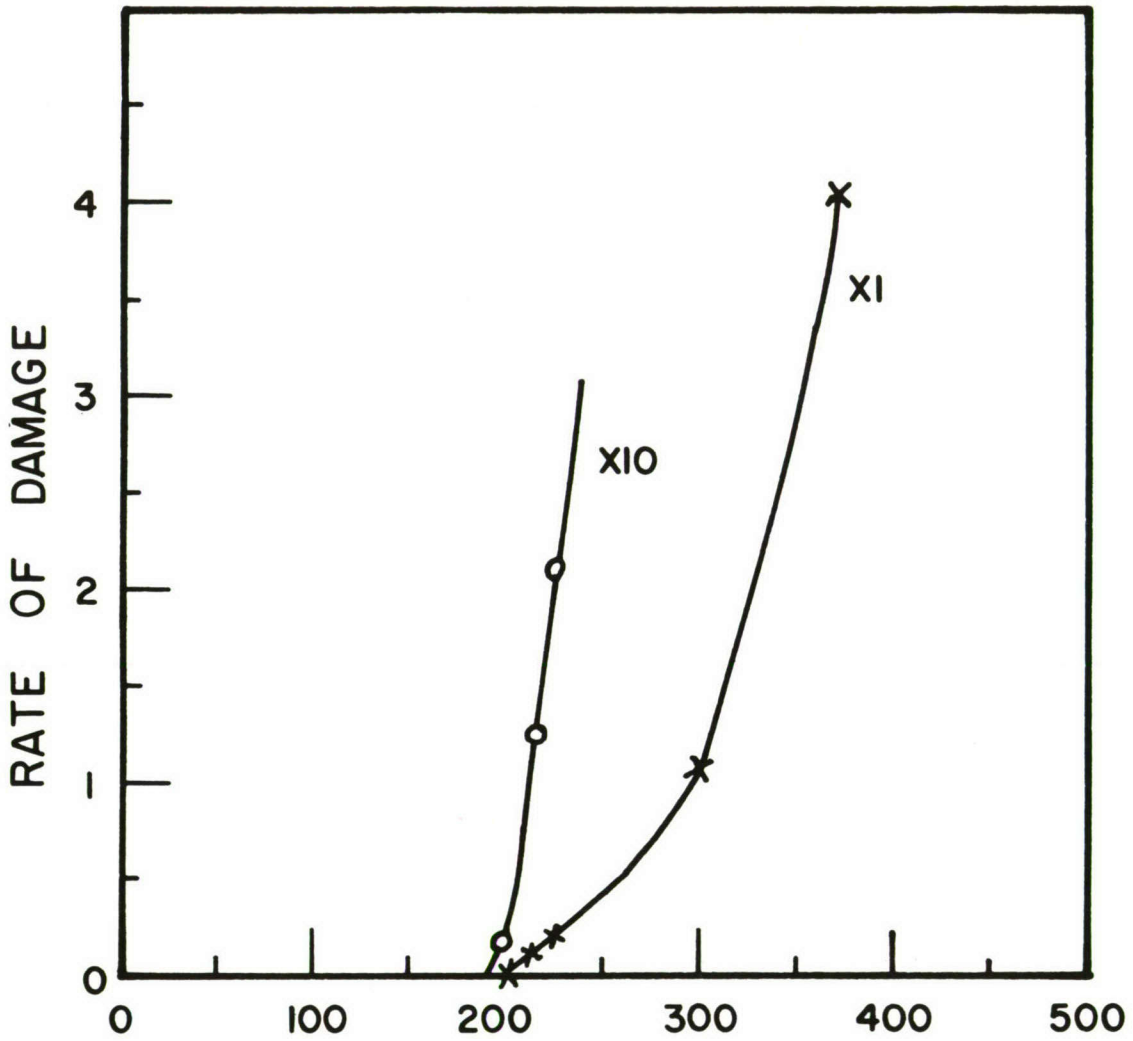

RATE OF DAMAGE
0
1
2
3
4
0
100
200
300
400
500
X10
X1
ELECTRON ENERGY, keV

FIG. 7

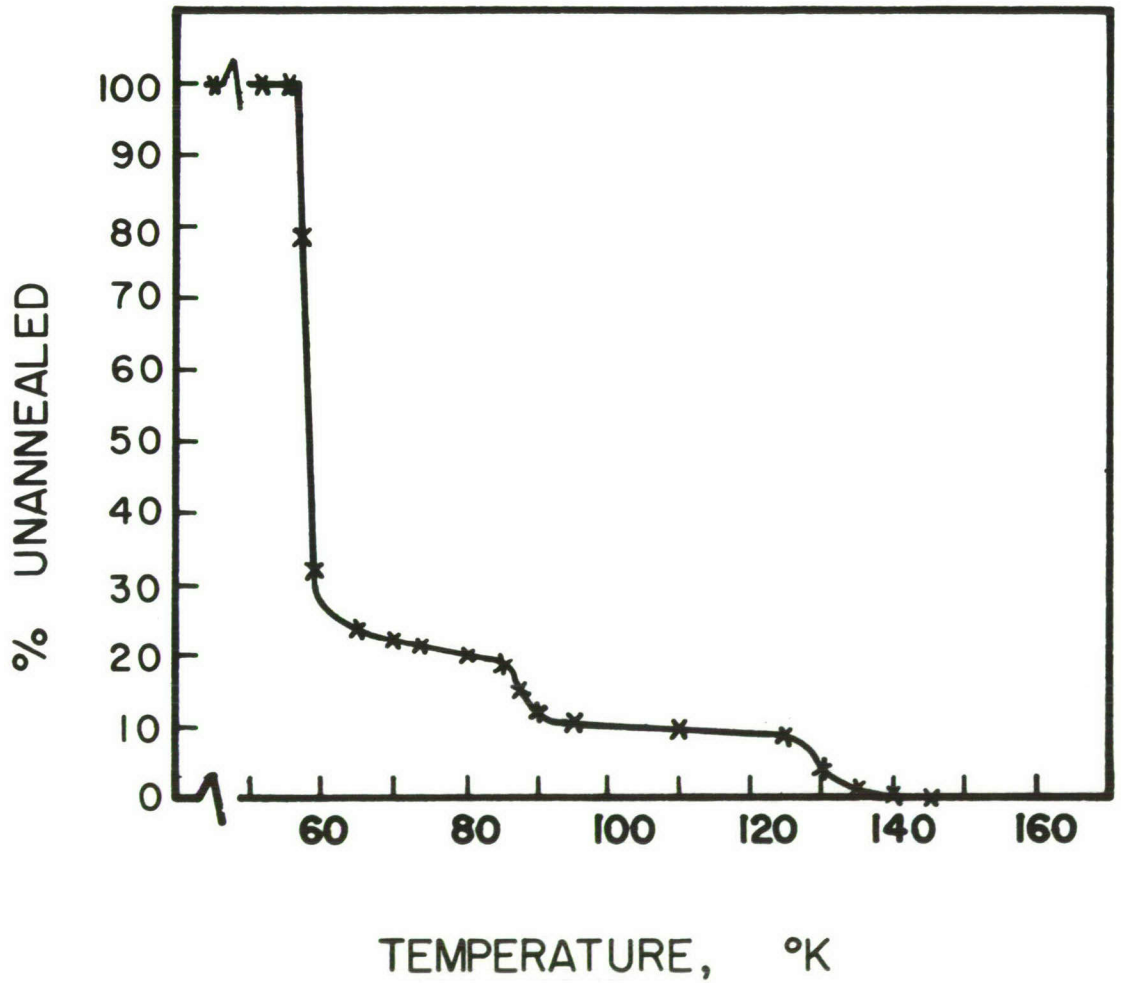

% UNANNEALED
100
90
80
70
60
50
40
30
20
10
0
60
80
100
120
140
160
TEMPERATURE, °K

FIGURE 8

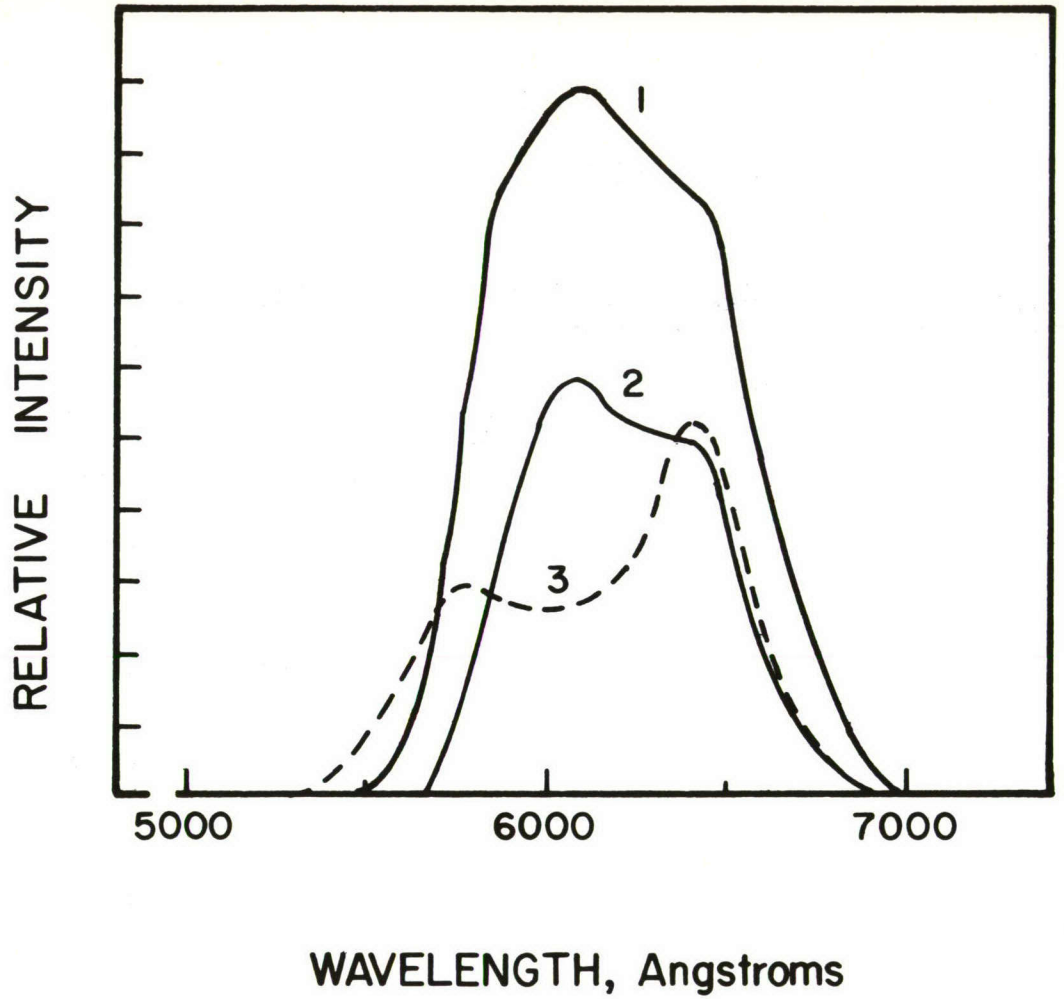

RELATIVE INTENSITY
1
2
3
5000
6000
7000
WAVELENGTH, Angstroms

FIGURE 9

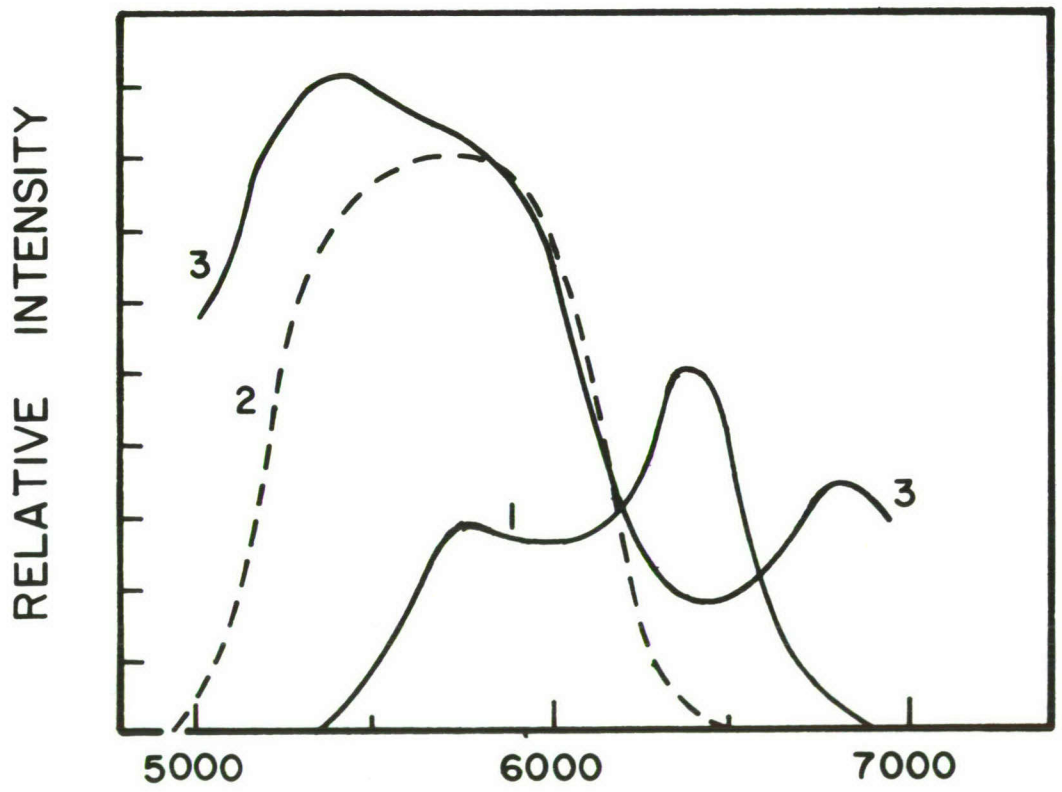

RELATIVE INTENSITY
3
2
1
3
5000
6000
7000
WAVELENGTH, Angstroms

FIGURE 10

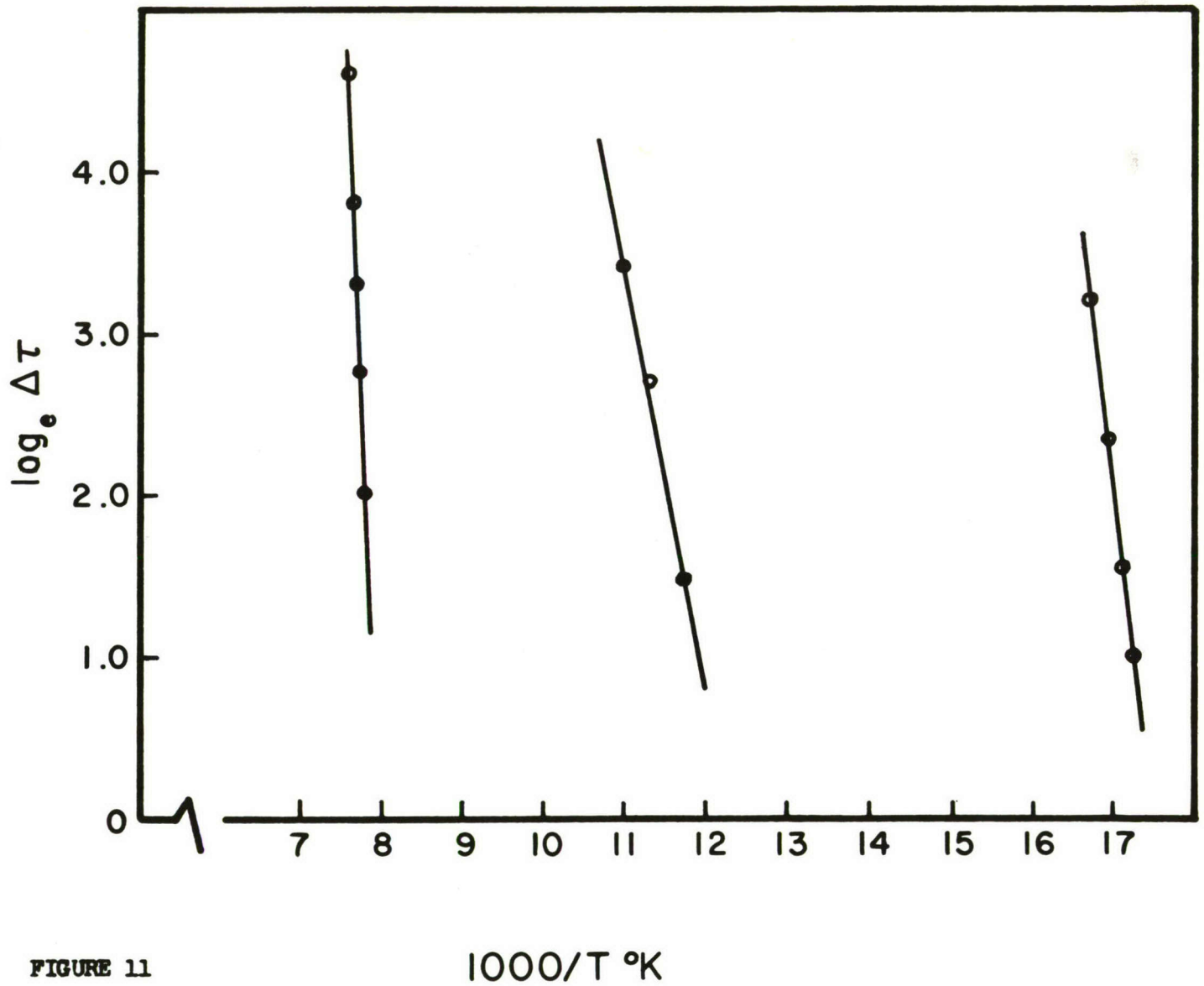


FIGURE 11

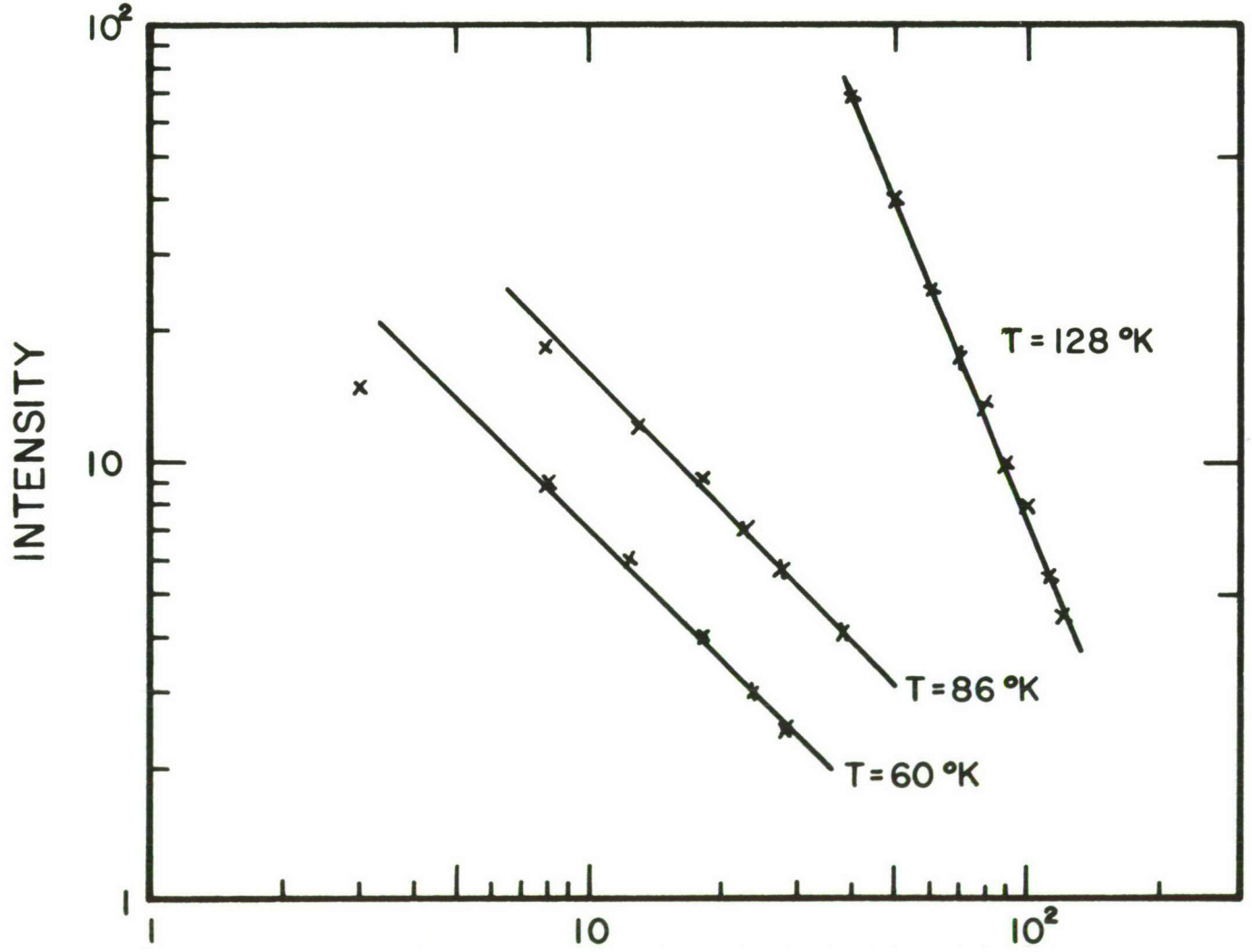


FIGURE 12

(U) MODIFICATION OF THE EFFECTS OF ACUTE LETHAL DOSES OF IONIZING RADIATION IN THE RHESUS PRIMATE[1]

by

George S. Melville, Jr.
George W. Harrison, Jr.
Robert J. Young

[1]This work was accomplished at the USAF School of Aerospace Medicine and at the Radiobiological Laboratory of the University of Texas and the United States Air Force and was supported, in part, by funds from Air Force Contracts 41(657)-382 and 41(609)-2005.

25 June 1964

CERTIFICATE

I certify that the experiments discussed in the paper, "Modification of the Effects of Acute Lethal Doses of Ionizing Radiation in the Rhesus Primates," by Capt George S. Melville, Jr. et al, were conducted according to the "Principles of Laboratory Animal Care" of the National Society for Medical Research.

Dale D. Boyd

DALE D. BOYD
Major, USAF, VC

(U) MODIFICATION OF THE EFFECTS OF ACUTE LETHAL DOSES OF IONIZING RADIATION IN THE RHESUS PRIMATE

INTRODUCTION

The experiment to be described herein represents an effort to advantageously combine three different approaches to the problem of radiation protection into a system which will enable a primate to survive an acute, otherwise lethal dose of roentgen rays.

In the radiation dose range between 700 and 1000 r in the Macaca mulatta rhesus monkey, two factors contribute to the death of the animals from irradiation: 1) destruction of hematopoietic tissue as exemplified by the bone marrow, and 2) functional failure of the gastrointestinal tract. The pathogenesis of acute radiation sickness has been described (5) and Quastler (23) among others, has observed the nature of the intestinal damage. The total treatment is designed to cope with these facets and their sequelae.

A homogeneous mixture of two thiols - S,β-aminoethylisothiuronium dibromide (AET) and cysteine hydrochloride - known to protect rats (14) and monkeys (17, 30) has been given orally and intravenously prior to irradiation. AET has been shown to protect the intestinal tract (10) and more particularly, lymphoid tissue and bone marrow (29). Cysteine is a very effective radioprotector when given intravenously (28).

The basic thiol mixture has been augmented by two radioprotective agents which very probably work through their pharmacologic properties to lower tissue oxygen tension. 5-hydroxytryptamine creatinine sulfate (serotonin)

NO OBJECTION TO PUBLICATION ON GROUNDS OF MILITARY SECURITY OFFICE OF INFORMATION AEROSPACE MEDICAL DIVISION
1 Jul 64 R. J. Kenny

is a potent vasoconstrictor which is an excellent protector in small mammals (9). Semenov (25) demonstrated that cholinergic drugs potentiate the action of serotonin. We have used acetylcholine in the rat (13) and methacholine in the monkey (16) for this purpose.

It has been observed in the monkey (12, 16) that even in protected animals surviving lethal radiation leukopenia is not prevented. Also at a dose such as 900 r, the intestinal epithelium is damaged and denuded (32). Thus infection and dehydration and other symptoms can and do occur. Over a 6-year period, a regimen of antibiotic,fluid, and other symptomatic therapy has been evolved in these laboratories (15, 16) and was adhered to in this experiment.

We have chosen transplantation of autologous marrow to afford living cells to repopulate damaged marrow sites. In the marrow-treated monkey, regeneration seems to occur three to five days earlier than in untreated or chemically-protected animals (20). Autologous marrow infusion has improved survival in x-irradiated dogs (1) and monkeys (2). Further, AET and antibiotics and bone marrow have been successfully combined in mice to afford survival for 60 days at 2600 r of gamma rays (4). Such a technique has been proven feasible in the monkey (31). The use of autologous bone marrow eliminates the problem of encountering the secondary disease syndrome described in monkeys by De Vries (6). This foreign bone marrow reaction occurs whenever homografts are utilized and is considered to be an immune response of the graft to the host.

MATERIALS AND METHODS

The primates used in this study were Macaca mulatta rhesus monkeys, weighing between 5 and 8 pounds and ranging in age from 30 to 40 months. They were selected on the basis of prior experiments which indicated that animals of this age and weight had sufficient marrow remaining in their extremities, and that the fatty marrow replacement seen in older animals had not occurred to any great extent.

Four groups of 6 animals each were designated for concurrent studies: a control group (C), a bone marrow group called bone marrow II (BM II), a chemicals group (Ch), and a chemicals plus bone marrow group designated I (Ch + BM I). Two months subsequent to the main experiment, a second chemicals plus bone marrow group was done and designated II (Ch + BM II). This second combined treatment group received a revised chemical treatment which utilized the results of Ch + BM I. Comparisons have also been made with an equally important group of six monkeys which received autologous bone marrow in a previous experiment (35); these animals have been designated bone marrow I (BM I). The so-called 'control' group were animals which were irradiated and then received only postirradiation symptomatic therapy.

All the animals were exposed to 900 r of x-rays and received a sedative prior to irradiation. None of the animals were under sedation during radiation exposure. With the exception of the C and Ch groups, all the other animals received intravenous autologous marrow within an hour after radiation exposure. The marrow was withdrawn just preceding chemical treatment and/or irradiation.

Bacteriologic and hematologic baseline determinations were performed during a 2-week period immediately preceding irradiation. Temperature, weight, and water intake values were also established prior to irradiation.

Irradiation:

The 900 r whole-body irradiation received by all six groups was delivered at the rate of 17.3 r/min. ± 1 r, from a Picker x-ray machine, 250 kvp, 18 milliamperes (ma) with a 1 mm. aluminum and 0.25 copper filtration. During irradiation the monkeys were restrained in a cylindrical exposure tube, approximately 8 inches in diameter and 16 inches tall, the sides of which are 0.25 inch aluminum rods on 2-inch centers, inserted in circular lucite discs 0.75 inches thick, which comprise the top and bottom. The exposure tube was rotated in the vertical plane at 3.5 revolutions per minute during exposure. The target distance to the center of the cage was 110 cm. The radiation dose was measured by means of 25 r and 100 r Victoreen R meters. Readings on these meters were corroborated by a Victoreen Rate Meter and by chemical dosimetric measurements. The x-ray machine was monitored during all the exposure times to insure that the stated settings were maintained and the exposure times calculated from the measured dose rates were adhered to.

Chemical Treatment:

The radioprotective agents used in this experiment were Mecholyl (methacholine bromide, Merck, Sharp & Dohme), AET (2-aminoethyl-isothiuronium dibromide, prepared by the method of Shapiro (26), cysteine (l-cysteine HCL.H_2O,Calbiochem) and 5-HT (serotonin creatinine sulfate, Calbiochem). Table I gives the body weight doses, time of administration prior to irradiation, total volume administered, and the buffer used. The sedative was

administered before any of the other treatments except where noted in Table I. Mecholyl was administered orally by tablet. AET and cysteine were given together in a homogeneous aqueous solution by gavage and intravenously and each animal received material by both routes as shown in Table I. Doses were prepared by neutralizing the agents with 2.5 N NaOH to pH 7.3 so that transformation of AET to the form of mercaptoethylguanidine (MEG) could be effected. They were then brought to volume with the appropriate buffer solution. 5-HT was dissolved in distilled water, no pH adjustment was made; the aqueous solution was given intraperitoneally.

Bone Marrow Treatment:

Animals were anesthetized with 10.5 mg/lb. phenobarbital sodium or other drugs (see discussion). An incision was made through the skin and underlying tissue at the distal ends of the humeri, tibiae, and radii. The bony cortex was exposed by scraping away the periosteum, and a size 43 bit utilized to bore an opening into the medullary cavity. Into this opening was inserted a 13 gauge needle, the tip of which had been removed 2 inches from the hub and the shaft threaded to facilitate insertion. (Fig. 1) A 3-foot long section of rubber tubing containing a two-way stopcock was attached to the needle at one end and the collection flask at the other. In the middle of the tubing a cylindrical aluminum housing was interposed. This housing could be unscrewed to permit the insertion of a 30-mesh stainless steel disc to screen the marrow.

A 50 ml. test tube was inserted in the 500 ml. collecting flask; the rubber tubing was attached to a glass tube which is bent at a 90° angle and inserted through a rubber stopper into the test tube (Fig. 2). Suction was provided by means of a rubber tubing which was attached to the side area of the suction flask and then to a 1/2 H.P. rotary suction pump.

The test tube was primed initially with 4 ml. Gey's BSS plus 1 ml. heparin. Following the aspiration of marrow from each bone, a like amount of Gey's solution and heparin was aspirated through the tube by immersing the needle into a small pyrex dish which contained the solution. This cleaned the tube and assisted in screening the marrow.

The collected cells were centrifuged at 2,000 rpm for 20 minutes, supernatant decanted, leaving a total volume of approximately 12 to 15 ml. of packed cells plus fluid, then stored in a refrigerator at 4° C. until time of injection. Generally 3×10^8 to 8×10^8 nucleated cells were harvested.

Postirradiation Clinical Treatment:

The postirradiation therapy was entirely based upon the treatment of symptoms. In general, anoxeria and dehydration were treated by the use of intravenous fluids containing nutrients and vitamins. Transfusions — either of plasma or whole blood — were not used except in very rare instances in the later stages of the 30-day period. Moribund or severely debilitated animals were given fluids and stimulants. When clinical evidence indicated the presence of systemic infection, broad-spectrum antibiotics were given while blood and stool cultures underwent microbiologic examination and sensitivity disc results suggested a specific drug for use. Water intake, urine output, and urine pH + specific gravity were noted and utilized to suggest treatment. The regular diet was augmented with fruit and potatoes; animals were given every opportunity to increase their food intake if they so desired. When severe hypothermia occurred animals were placed in an incubator. The principles and practice of primate care in our colony has been described by Gisler (7).

Biochemistry and Hematology:

Peripheral blood sampling was performed in accordance with the standard technique which has been described by Young, et al (35).

Serum glutamic pyruvic transaminase determinations were performed by the method of Steinberg, Baldwin and Astraw (27). SGPT determinations were performed in animals surviving in excess of 30 days.

Urea nitrogen determinations were performed utilizing the Urograph* as the diagnostic tool. Electrolyte values were determined with a flame photometer according to the method of Proehl and Nelson (22). Serum chloride determinations were performed according to the method of Schales and Schales (24). Total protein was determined according to Kingsley (8).

The A/G ratio was elicited by serum protein analysis utilizing a paper Electrophoresis System, developed by Beckman Instruments, Inc. (11).

Blood cell counts for erythrocytes, platelets, reticulocytes, leukocytes, lymphocytes, and polymorphonuclear neutrophils were performed according to Wintrobe (34). Hematocrits were determined by the microhematocrit method in duplicate; reticulocytes were stained by Brecher's methylene blue method (3) Hemoglobins were done by the Cyanmethemoglobin method (30), using an Acuglobin Hemoglobin Standard.

---

*Warner-Chilcott, Morris Plains, New Jersey

RESULTS

Mortality:

The survival of the animals in this experiment is compared to that of those in Bone Marrow I and Chemicals + Bone Marrow II in figure 3. At 30 days postexposure only one animal from BM II is alive, whereas three animals were alive at this point in BM I (35). Mean survival times are 13 and 12 days, respectively, for the animals in these groups which die before 30 days. Only three animals in CH + BM I are alive at 30 days post-exposure compared to six survivors in CH + BM II. Those animals dying in less than 30 days had a mean survival time of 12 days. Two monkeys survived for 30 days in the Chemicals group; the four which expired has a mean survival time of 9.5 days. All of the Control group were dead by day 30 with a mean survival time of 10 days.

At 360 days postexposure none of the animals in the bone marrow groups were alive. Death for these monkeys took place from 119 to 155 days. There is one survivor in the Chemicals group at 360 days; the other animal died on day 119. Only one death has occurred in the chemicals plus bone marrow monkeys that survived 30 days, a Ch + BM II animal that died on day 53.

Clinical Observations:

The clinical symptoms observed, the day of death, and the numbers of bone marrow cells transfused are shown for each animal arranged in order of increasing survival in tables II, III, and IV. Loss of weight occurred to a marked degree in the Control group monkeys and the BM II animals, but

was less pronounced in all other groups. Anorexia was observed in moderate to severe form in all of the groups but BM I in which the monkeys were only slight affected. There was also a lower incidence of diarrhea in the Bone Marrow I group than in any other group. The Control and BM II monkeys were the most severely affected with regard to diarrhea. Bloody diarrhea was manifested only in the Control group. Febrile responses occurred to a lesser extent in control and bone marrow animals than in those of the other groups. In many cases, including monkeys which received marrow, hypothermia was observed, especially during the days just before exitus.

In general, there appears to be very little diminution of clinical symptoms in the survivors as compared to the non-survivors. The most notable exception is monkey C45 in the Ch + BM I group; this animal demonstrated less severe symptomology in all categories than any other animal in its group. In the Chemicals group, the occurrence of diarrhea is lower in the survivors.

Hematology:

Figure 4 is a hemogram showing the mean cell counts by experimental groups for total white cells, absolute neutrophils, absolute lymphocytes, and platelets for those animals that survive through day 50 postexposure. White blood cell counts fall below 1,000 cells per cc. of blood by the end of the first week. By day 15 all but the Chemicals group survivors are showing increased numbers of white cells. Only the monkeys in Ch + BM II are recovered by day 30. Counts for the Ch survivors were essentially normal by day 37. Bone marrow I survivors had reached normal levels for white cells by day 56 (35), but the poorest recovery is seen in BM II

where the value is only 40% of baseline values by day 50. The Ch + BM I survivors had reached 70% of baseline by day 50. As may be seen in the plots for group mean absolute neutrophil counts, this cell type seems to be mainly responsible for the recovery patterns seen in the total white blood cells.

Mean platelet counts for all of the groups fall below 200,000 cells per cc. of blood by the end of the first week postexposure. Survivors in BM I + Ch BM II groups begin to recover by the end of the second week. The platelet counts for the Ch group are still falling at the end of the second week but have risen to a point similar to that of most of the other groups by the end of the third week. Platelet values are within the normal range by day 30 for the survivors in the Ch and Ch + BM I groups. The slowest recovery is shown by the lone survivor in the BM II group.

Figure 5 is a hemogram showing the mean cell counts for groups against time postexposure for red blood cells, reticulocytes, hemoglobin, and hematocrit for those monkeys surviving through day 50 postirradiation. The responses of erythrocyte counts, hemoglobin, and hematocrit are similar in nature. All of these values are depressed less severely in the BM I and Ch + BM II groups than in the others, while the Ch group is depressed most severely. The counts return to normal in all animals by day 50 with the exception of the BM II survivor who is only at 70% of baseline by this time.

The numbers of reticulocytes seen in the peripheral blood are lower than baseline levels in all groups by the end of the first week post-exposure but are least depressed in the monkeys in Ch + BM I. By day 15 only the survivors in the Ch group show no evidence of a return to normal counts. These animals then experience an extreme reticulocytosis which

lasts until about day 37. A less marked reticulocytosis is noted in the BM I and Ch + BM I groups by day 22.

Clinical Chemistries:

The results of the Biuret determination of total serum protein is shown in Table V. The average baseline values of serum total protein for all of the groups indicate a range of 7 to 8.5 grams per 100 ml of serum. All of the animals which did not receive both chemicals and bone marrow show decreased amounts of total protein. There is a less consistent trend in Ch + BM I and Ch + BM II which is transient and has returned to normal by day 30. The results of electrophoresis determinations are shown in Tables VI and VII. The A/G ratios mirror the pattern shown by total protein and indicate that the problem is probably that of an albumin loss which corrects itself by day 30 in most surviving animals. In Ch + BM II only one animal—C34—is notably different from his preirradiation mean.

Table VIII shows the results of blood urea nitrogen determinations upon these primates. There is no discernible pattern in these values and a majority of the changes are transient in nature. The decreases when they do occur are observed between days 8 and 22.

Table IX shows the values measured for blood potassium. There are very few marked changes. The response seems to depend only upon the fact of irradiation and when seen is a transient decrease during the first 15 days postexposure; many animals actually equal or exceed their baseline averages by the fourth week after radiation.

Table X shows the results of blood sodium determinations. With the exception of two monkeys in Ch + BM II which show a marked increase during

the 4th week postirradiation, all other results are normal and show no patterns or trends.

Table XI describes the chloride determination results. There are no changes beyond the normal fluctuations in Ch + BM II; animal C34 shows a marked drop on day 30. In the other groups, the behavior of the values are not consistent; however, there are more depressions than elevations during the first 15 days in the 24 animals. Since the normal range is 115 milliequivalents plus or minus 10%, none of the transient changes purport clinical significance.

An approximate "water balance" was arrived at by calculating the difference between the volume of drinking water consumed each day and the volume of water calculated from the 24-hour urine sample using the measured specific gravity of the sample. Figure 6 shows this difference of water intake and water output averaged every three days for each treatment group and plotted against postirradiation time. Baseline differences for all groups were positive; after exposure most differences become negative with the exception of Ch + BM II. Since this group of animals ultimately survived the best, this result indicates the potential importance of maintaining water consumption at a level above that excreted as urine.

DISCUSSION

It would not be factual to intimate that the procedures utilized for Ch + BM II were the result of serendipity. Sixteen other monkeys were studied under various experimental conditions — nearly all of these died within a few hours after exposure from toxic complications.

The problem of drug toxicities, immediate radiation effects, and surgical trauma proved extremely troublesome since in all cases these three appeared to be additive. It must be emphasized that in order to arrive at useful drug dose levels for a combination treatment, the entire procedure including radiation should be examined in the experimental animal of choice.

Sodium Pentobarbital or Sodium Pentathal had been used in our surgery for several years and Sodium Pentobarbital as Nembutal had been part of a successful radioprotective system (17, 18). It became obvious, regardless of other toxic problems, that Nembutal was not well-tolerated by the monkeys after a radioprotective dose of mecholyl. Therefore, in Ch + BM I, the animals completed the surgical procedure and were awake before any chemicals were given with the exception of C45. This monkey was the only animal of seven receiving mecholyl prior to surgery which did not die an early toxic death. As seen in Table I the dose of mecholyl was also reduced.

The drugs in this system tend to depress and sedate the monkey when used alone and the effect seems to be additive when they are combined. Reduction of the amounts of the thiol mixtures did not help nor did the substitution of Trimar for Nembutal during surgery. This led us to believe

that the serotonin dose was the more important toxic problem. Under the conditions of this experiment, the 15 mg./kg. body weight was given intramuscularly instead of intraperitoneally but was tolerated no better by the monkeys. Using the levels for the mecholyl and the MEG-Cysteine dose shown in Table I for the Ch + BM II group, the serotonin dose was reduced stepwise until no toxic symptoms were noted at 6.5 mg./kg. body weight. This system was then combined with the surgical technique using ether and was subsequently used to do the Ch + BM II group.

It will be noted that the oral administration of the thiol mixture was pushed back to one hour prior to irradiation for the final experiment. Such a time is in accord with our studies in the rat (18) and with other concurrent studies in the primate (16). It seemed much more logical, indeed, for the oral dose to be administered before the intravenous dose—gastrointestinal absorption should be considerably slower.

Despite the fact that the clinical opinion of those providing symptomatic treatment to heavily-irradiated primates that fluid and electrolyte therapy were essential to survival, very few changes were observed in sodium, potassium, and chloride on this sampling schedule. Polikarpova (21) has reported a twofold increase in serum chloride during the first 24 hours after the exposure of young rhesus primates to 700 r of x-rays; however, the values were below normal by day 5. Our findings are supported by those of Wilson (33) whose data indicated that sodium, potassium, and chloride were normal by 12 days after 725 r of Co^{60} gamma rays.

The "water balance" data delineate the groups much more sharply and this separation corresponds to the survival results. These data, of course, mainly observe the survivors who may be presumed to be less sick; on the

other hand, the sickest animals received the most therapy. Recovery processes probably call upon all available energy sources including that from the endogenous reserve in the fat depots. The release of energy from this source would produce metabolic water which would not be measured — only the exogenous source was measured — but could result in increased urine output which would create an apparent negative "water balance". Thus the animals in Ch + BM II may not only be in better condition with regard to hydration and electrolytes but also may not be calling as heavily upon their endogenous energy reserves.

It is easy to determine that the two groups of animals which received autologous bone marrow only are somewhat different in their response. This is probably to be expected since the monkeys will respond as individuals and there may well be a spectrum of degrees of "take" of the graft. In this heterologous population of primates, there is no reason to assume that the animals are precisely comparable even though they were matched with respect to the obvious criteria.

There would appear to be little question that — at 900 r — the entire treatment is necessary to achieve a high order of indefinite survival. In some 15% of chemically-protected monkeys surviving lethal doses of x-rays for more than 30 days, Pitcock (19) has described pathologic evidence for a subacute death process or a limitation of the recovery process, if you will, which results in death at about 100 days postirradiation; bone marrow infused monkeys in this experiment and previous studies (24) seem to face the same limitation. It is still to be definitely proven if either part of the combined system contributes more to the long-term survival than the other.

The combined treatment offers a considerable reduction in the amounts of protective material which must be administered for an effect. In other protection studies in the rhesus primate at 800 r of x-rays (12), almost 2.3 grams per kilogram body weight of the MEG-Cysteine Thiol mixture was required to save only half of the animals irradiated. For Ch + BM II, barely one gram per kilogram body weight of all drugs was administered prior to irradiation for better survival at a higher radiation dose. The opportunity to give less material is important to both the pharmacologist and the pharmacist when one looks ahead to toxic side effects and to problems concerning dosage forms which might be encountered in the human primate.

ACKNOWLEDGEMENTS

The authors wish to particularly acknowledge the contributions to this research by the following individuals: Major D. D. Boyd, Mr. W. Fowler, Lt. Col. M. D. Harris, Mr. T. P. Leffingwell, Mrs. B. Olberding, TSgt C. Pendley, Dr. G. R. Vela, and Capt. T. Wolfle.

REFERENCES

1. Alpen, E. L. and S. J. Baum. Modification of x-radiation lethality by autologous marrow infusion in dogs. Blood. 13:1168-1175 (1959).

2. Ambrus, C. M., E. T. Felts, and J. W. Byron. Autologous Bone Marrow and Lymph Infusions as Aids in Radiation and Chemotherapy. Proc. AM. Assoc. Cancer Res. 3:2 (1959).

3. Brecher, G. New Methylene Blue as a Reticulocyte Stain. Am J. Clin. Path. 19:895-896 (1949).

4. Burnett, W. T., Jr., and D. G. Doherty. Additive effect of S, β-aminoethylisothiuronium.Br-HBr, Bone Marrow, and Streptomycin on Gamma-irradiated Mice. Radiation Research. 3:217 (1955).

5. Cronkite, E. P. and G. Brecher. Radioactivity. Effects of whole-body irradiation. Ann. Rev. Med. 3:193-214 (1952).

6. DeVries, M. J., B. G. Crouch, L. M. Van Putten, and D. W. Van Bekkum. Pathologic Changes in Irradiated Monkeys Treated with Bone Marrow. J. Nat. Cancer Inst. 27:67-97 (1961).

7. Gisler, D. B., R. E. Benson, and R. J. Young. Colony Husbandry of Research Monkeys Annals. New York Academy of Science. 85:758-768 (1960).

8. Kingsley, G. R. The Direct Biuret Method for the Determination of Serum Proteins as applied to Photoelectric and Visual Colorimetry. J. Lab. and Clin. Med. 27:840-845 (1942).

9. Langendorff, H. and R. Koch. Untersuchungen uber einen biologischer Strahlenschutz. XVIII. Die Wirkung zentralerregender Pharmaka auf das bestrahlte Tier. Strahlentherapie. 102:58 (1957).

10. Maisin, J. R., Novelli, G. D., D. G. Doherty, and C. C. Congdon. Chemical Protection of the Alimentary Tract of Whole-body X-irradiated Mice. I. Changes in weight, histology, and cell division in relation to nucleic acid and protein synthesis. Int. J. Rad. Biol. 2:281 (1960).

11. Manual: Spinco Model R., Paper. Electrophoresis System, Operating Instructions. Belmont, Calif., Spinco Div., Beckman Instruments, Inc. (1955).

12. Melville, G. S., Jr., R. E. Benson, T. P. Leffingwell, and G. W. Harrison, Jr. Radioprotection in Primates: A Preliminary Report. SAM-TDR 61-3. (January 1961).

13. Melville, G. S., Jr., G. W. Harrison, Jr., and K. M. Gallo. Chemical Radioprotection. Aerospace Medicine. 34(10):938-941 (1963).

14. Melville, G. S., Jr., G. W. Harrison, Jr., and T. P. Leffingwell. Unpublished data.

15. Melville, G. S., Jr., G. W. Harrison, Jr., T. P. Leffingwell, W. H. Whitcomb, and R. J. Young. Protection against Ionizing Radiation: x-irradiated monkeys receiving preirradiation prophlayxis and postirradiation therapy. SAM-TDR 62-103 (December 1962).

16. Melville, G. S., Jr., G. W. Harrison, Jr., A. A. McDowell, and T. P. Leffingwell. Chemical Protection in Irradiation Monkeys. Proceedings, International Symposium on Bone Marrow Therapy and Protection in Irradiated Primates, August 15-18, 1962, Rijswijk (Z. H.), The Netherlands.

17. Melville, G. S., Jr., G. W. Harrison, Jr., and T. P. Leffingwell. Radioprotection with AET-Cysteine in the Rhesus Monkey. Radiation Res. (in press).

18. Melville, G. S., Jr., and T. P. Leffingwell. Potentiation by Pentobarbital of the Protective Effect of AET against Gamma Rays. USAF School of Aerospace Medicine TDR 61-87.

19. Pitcock, J. A. and G. S. Melville, Jr. Subacute Radiation Death in Chemically-Protected Monkeys. Rad.Res. 16:692-698 (1962).

20. Pitcock, J. A., W. H. Whitcomb, R. J. Young, G. S. Melville, Jr., and D. R. Anderson. Histopathologic Effects of Homologous Bone Marrow Transfusions in Irradiated Monkeys (Macaca Mulatta). To be published.

21. Polikarpova, L. I. and A. Ya. Shulyatikova. Chloride Content of the Blood of Rhesus Monkeys after Irradiation. Meditsinskaya Radiologiya, 3:79-80 (1961).

22. Proehl, E. C. and W. P. Nelson. The Flame Photometer in Determination of Sodium and Potassium. Amer. J. Clin. Pathol. 20:806-813 (1950).

23. Quastler, H. The Nature of Intestinal Radiation Death. Radiation Res. 4:303-320 (1956).

24. Schales, O. and S. S. Schales. A Simple and Accurate Method for the Determination of Chlorides in Biological Fluids. J. Biol. Chem. 140:879-884 (1941).

25. Semenov, L. F. Assay of Indoleamine Compounds in the Prevention of Radiation Sickness. Med. Radiol. 5(5):47 (1960).

26. Shapira, R., D. G. Doherty, and W. T. Burnett, Jr. Chemical Protection against Ionizing Radiation: III. Mercaptoalkylguanidine and related isothiuronium compounds with protective activity. Rad.Res. 7:22-34 (1957).

27. Steinberg, D., D. Baldwin, and B. H. Ostrow. Clinical Method for Assay of Serum Glutamic-oxalacetic transaminase. J. Lab. and Clin. Med. 48:144-151 (1956).

28. Thomson, J. F. Radiation Protection in Mammals. Reinhold Publishing Co. Ch 5 (1962).

29. Urso, P., C. C. Congdon, D. G. Doherty, and R. Shapira. Effect of Chemical Protection and Bone Marrow Treatment on Radiation Injury in Mice. Blood. 13:665 (1958).

30. Van Lancker, J. L., R. C. Wolf, and J. B. Mowbray. Protection of primates against lethal doses of x-irradiation. Nature. 194:492-493 (1962).

31. Whitcomb, W. H., R. J. Young, G. S. Melville, Jr., J. A. Pitcock, and D. R. Anderson. Chemical Protection of Irradiated Monkeys in Combination with Autologous and Homologous Marrow Transfusion. To be published.

32. Wilson, S. G. Radiation-Induced Gastrointestinal Death in the Monkey. SAM-TDR 59-77.

33. Wilson, S. G., W. H. Whitcomb, and D. R. Anderson. The Sequential Development of Lesions in Fatally Irradiated Monkeys. Semi-Annual Progress Report, The Radiobiological Laboratory of the University of Texas and the USAF, 1 April 1959 to September 1959, AF Contract 41(657)-149.

34. Wintrobe, M. M. The Principles and Technic of Blood Examination. Clinical Hematology. Philadelphia, Pa.: Lea and Febiger, 5th ed., (1961).

35. Young, R. J., W. H. Whitcomb, G. S. Melville, Jr., and D. R. Anderson. Radioprotection in Primates. Proceedings, International Symposium on Bone Marrow Therapy and Protection in Irradiated Primates, August 15-18, 1962, Rijswijk (Z. H.), The Netherlands.

Table I

Chemicals Received By the Irradiated Monkeys

Treatment	Chemicals Only	Chemicals + Marrow I	Chemicals + Marrow II
Mecholyl orally	100 mg/kg, 130 min. prior#	87 mg/kg, 60 min. prior*	35 mg/kg, 90 min. prior
AET-cysteine intravenously	60-450 mg/kg, 30 min. prior, 9.5 ml total volume, H_2O	60-450 mg/kg, 30 min. prior, 8.5 ml total volume H_2O	48-360 mg/kg, 30 min. prior 8 ml total volume, H_2O
AET-cysteine orally	190-1140 mg/kg, 20 min. prior, 15 ml total volume, H_2O	190-1140 mg/kg, 20 min. prior, 15 ml total volume, PO_4 buffer	90-450 mg/kg, 60 min. prior, 10 ml total volume, PO_4 buffer
5-HT intra-peritoneally	15 mg/kg, 20 min. prior, 5 ml total volume	15 mg/kg, 20 min. prior, 5 ml total volume	6.5 mg/kg, 20 min. prior 1.5 ml total volume

\# Dose given before sedation

* One animal (C45) given dose before sedation (230 min. prior)

\+ 1 ml of procaine included as part of the dose except for one animal (B52)

TABLE II

Clinical Summary for the Irradiated Controls and Chemically-Treated Monkeys

Animal	Day of Death	Weight Loss	Anorexia	Diarrhea	Febrile Response
Control Gp.					
C49	5	++	+++	+++	None
C54	7	+	+++	+++	None
D87	11	++	+++	+++	None
10H	12	+++	+++	+++	+
E10	12	++++	++++	++++	None
E69	13	++++	+++	++++	+
Chemical Gp.					
B20	1	N. R.	N. R.	N. R.	N. R.
B52	7	N. R.	N. R.	N. R.	N. R.
C83	15	++	+++	++	None
E12	15	+	+++	+++	None
E00	119	+	++	+	+++
C61	360 +	+	+++	None	+

TABLE III

Clinical Summary for Two Groups of Monkeys Receiving Autologous Marrow

Animal	Day of Death	# marrow cells x 10^8	Weight Loss	Anorexia	Diarrhea	Febrile Response
BM I Gp.						
84F	10	.57	++	N. R.	N. R.	None
30F	12	.35	+	++	++	+
118F	13	.32	+	++	+	++
106F	119	.68	+	+	+	None
98F	123	2.18	+	+	+	+++
127F	155	.48	++	+	+	None
BM II Gp.						
107E	6	3.15	+++	++	+++	+
D63	10	4.68	+	++++	+++	+++
B54	11	2.25	+	++++	+++	+
20K	12	.69	++	++	+++	None
D61	27	23.07	++	+++	+++	None
49H	119	8.67	+++	++	++++	None

TABLE IV

Clinical Summary for Two Groups of Animals Receiving the Combined Treatment

Animal	Day of Death	# marrow cells x 10^8	Weight Loss	Anorexia	Diarrhea	Febrile Response
Ch + BM I						
B40	6	4.43	N. R.	N. R.	N. R.	N. R.
B46	14	3.46	++	+++	++	+
B50	16	9.28	+	++	+++	++
B18	360 +	3.73	+	+++	++	+
B32	360 +	3.37	++	++	+++	+
C45	360 +	15.12	None	+	+	None
Ch + BM II						
B60	53	N. R.	+	+++	+++	+
B59	360 +	N. R.	+	++	++	+++
C24	360 +	N. R.	++	++	++	+++
C34	360 +	N. R.	+	+	+	++
C59	360 +	N. R.	+	++	++	+++
D50	360 +	N. R.	+	+	++	+

TABLE V

Total Protein as Grams of Protein per 100 ml. of Blood

Treatment and Animal #	Baseline Average	Days Postirradiation 8	15	22	30	37	44	50-56
Control								
C49	7.04	--	--	--	--	--	--	--
C54	7.66	--	--	--	--	--	--	--
D87	7.44	6.83	--	--	--	--	--	--
10H	7.84	7.49	--	--	--	--	--	--
E10	7.50	7.45	--	--	--	--	--	--
E69	8.08	6.77	--	--	--	--	--	--
Chemicals								
B20	7.70	--	--	--	--	--	--	--
B52	7.92	--	--	--	--	--	--	--
C83	7.50	6.68	--	--	--	--	--	--
E12	7.72	5.92	--	--	--	--	--	--
E00	8.18	8.10	5.37	5.93	5.70	8.06	5.85	7.45
C61	8.42	6.43	5.12	4.42	5.32	7.69	4.78	7.10
BM I								
84F	7.56	6.56	--	--	--			
30F	7.64	6.84	--	--	--			
118F	7.65	6.16	--	--	--			
106F	8.41	6.40	N. R.	5.16	7.37			
98F	8.43	7.08	6.64	4.00	7.21			
127F	8.59	7.08	6.64	6.56	7.37			
BM II								
107E	N. R.	--	--	--	--	--	--	--
D63	7.40	6.20	--	--	--	--	--	--
B54	N. R.	6.45	--	--	--	--	--	--
20K	7.06	6.59	--	--	--	--	--	--
D61	7.26	5.73	4.72	4.31	--	--	--	--
49H	7.71	7.07	5.45	6.30	5.93	6.93	6.13	6.13
Ch + BM I								
B40	7.13	--	--	--	--	--	--	--
B46	7.17	6.18	--	--	--	--	--	--
B50	6.95	6.22	4.95	--	--	--	--	--
B18	7.13	5.73	5.55	5.40	5.10	6.75	7.30	7.54
B32	7.00	6.22	6.46	6.32	6.20	7.63	7.05	7.20
C45	7.81	6.59	6.87	7.04	7.53	7.90	7.75	7.48
Ch + BM II								
B59	7.99	7.10	7.00	6.94	8.26			
B60	6.96	6.94	7.35	7.25	8.10			
C24	7.11	5.17	5.64	8.10	7.15			
C34	7.69	6.70	6.90	7.45	5.64			
C59	7.62	6.10	5.99	6.36	7.40			
D50	7.89	5.77	7.20	7.78	7.78			

TABLE VI
The Albumin to Globulin Ratio

Treatment and Animal #	Baseline Average	Days Postirradiation 8	15	22	30	37	44	50-56
Control								
C49	1.49	--	--	--	--	--	--	--
C54	1.38	--	--	--	--	--	--	--
D87	1.41	1.17	--	--	--	--	--	--
10H	1.43	.69	--	--	--	--	--	--
E10	1.36	.89	--	--	--	--	--	--
E69	1.36	.89	--	--	--	--	--	--
Chemicals								
B20	1.13	--	--	--	--	--	--	--
B52	1.78	--	--	--	--	--	--	--
C83	1.56	.85	--	--	--	--	--	--
E12	.98	.92	--	--	--	--	--	--
E00	1.00	.75	.59	.54	.61	.61	.61	.68
C61	1.68	1.04	.72	.69	.49	.86	.75	.75
BM I								
84F	1.34	1.13	--	--	--			
30F	1.20	1.31	--	--	--			
118F	1.42	1.70	--	--	--			
106F	1.46	1.86	N. R.	.87	1.68			
98F	1.52	1.49	1.00	.75	1.44			
127F	1.27	1.64	1.82	1.02	1.24			
BM II								
107E	N. R.	--	--	--	--	--	--	--
D63	1.42	.61	--	--	--	--	--	--
B54	N. R.	.67	--	--	--	--	--	--
20K	1.57	.96	--	--	--	--	--	--
D61	1.08	.54	.79	.54	--	--	--	--
49H	1.54	.89	.85	1.00	.85	.77	1.09	1.09
Ch + BM I								
B40	1.63	--	--	--	--	--	--	--
B46	2.23	1.13	--	--	--	--	--	--
B50	1.56	.85	.56	--	--	--	--	--
B18	1.78	.85	.71	.45	.50	.58	.76	.68
B32	1.27	1.04	1.17	.67	.96	.98	1.17	1.15
C45	.96	.92	.75	.79	.84	.89	1.03	.93
Ch + BM II								
B59	1.01	.82	1.22	.72	1.38			
B60	1.14	.92	.52	.64	1.04			
C24	1.15	.49	.56	.61	1.04			
C34	1.20	.81	.82	N. R.	1.22			
C59	1.14	.53	.59	.89	1.27			
D50	1.04	.96	.49	N. R.	.85			

TABLE VII

Albumin Expressed as Grams per 100 ml. of Blood

Treatment and Animal #	Baseline Average	Days Postirradiation 8	15	22	30	37	44	50-56
Control								
C49	4.18	--	--	--	--	--	--	--
C54	4.44	--	--	--	--	--	--	--
D87	4.36	3.69	--	--	--	--	--	--
10H	4.53	3.07	--	--	--	--	--	--
E10	4.32	3.50	--	--	--	--	--	--
E69	4.64	3.18	--	--	--	--	--	--
Chemicals								
B20	4.08	--	--	--	--	--	--	--
B52	5.07	--	--	--	--	--	--	--
C83	4.58	3.07	--	--	--	--	--	--
E12	3.82	2.84	--	--	--	--	--	--
E00	4.08	3.48	1.99	2.08	2.17	3.06	2.22	2.99
C61	5.30	3.28	2.15	1.81	1.76	3.54	2.06	3.05
BM II								
107E	N. R.	--	--	--	--	--	--	--
D63	4.32	2.35	--	--	--	--	--	--
B54	N. R.	2.59	--	--	--	--	--	--
20K	4.30	3.23	--	--	--	--	--	--
D61	3.76	2.00	2.08	1.51	--	--	--	--
49H	4.66	3.32	2.51	3.15	2.73	2.98	3.19	3.19
Ch + BM I								
B40	4.42	--	--	--	--	--	--	--
B46	4.95	3.28	--	--	--	--	--	--
B50	4.24	2.86	1.78	--	--	--	--	--
B18	4.56	2.64	2.28	1.67	1.69	2.50	3.14	3.09
B32	3.92	3.17	3.49	2.53	3.04	3.74	3.82	3.89
C45	3.86	3.16	2.95	3.10	3.46	3.71	3.95	3.59
Ch + BM II								
B59	4.00	3.20	3.85	2.91	4.79			
B60	3.69	3.33	2.50	2.83	4.13			
C24	3.73	1.71	2.03	3.08	3.65			
C34	4.19	3.02	3.11	N. R.	3.10			
C59	4.04	2.14	2.22	2.97	4.14			
D50	4.02	2.83	2.38	N. R.	3.58			

TABLE VIII

Blood Urea Nitrogen as Milligrams per 100 ml. of Blood

Treatment and Animal #	Baseline Average	Days Postirradiation 8	15	22	30	37	44	50-56
Control								
C49	28	--	--	--	--	--	--	--
C54	38	--	--	--	--	--	--	--
D87	28	20	--	--	--	--	--	--
10H	52	50	--	--	--	--	--	--
E10	32	35	--	--	--	--	--	--
E69	42	30	--	--	--	--	--	--
Chemicals								
B20	40	--	--	--	--	--	--	--
B52	40	--	--	--	--	--	--	--
C83	25	15	--	--	--	--	--	--
E12	40	15	--	--	--	--	--	--
E00	42	15	25	20	25	26	25	40
C61	45	20	20	25	15	22	33	25
BM II								
107E	N. R.	--	--	--	--	--	--	--
D63	28	25	--	--	--	--	--	--
B54	N. R.	40	--	--	--	--	--	--
20K	25	35	--	--	--	--	--	--
D61	25	25	15	15	--	--	--	--
49H	42	30	20	15	20	28	41	22
Ch + BM I								
B40	15	--	--	--	--	--	--	--
B46	55	40	--	--	--	--	--	--
B50	30	20	45					
B18	40	30	20	21	12	32	20	22
B32	30	20	10	10	25	34	35	35
C45	42	25	20	30	30	35	20	20
Ch + BM II								
B59	22	30	20	22	25			
B60	28	15	15	25	45			
C24	25	25	30	30	25			
C34	30	20	20	22	35			
C59	35	25	30	20	30			
D50	22	15	20	18	25			

TABLE IX

Potassium as milliequivalents per Liter of Serum

Treatment and Animal #	Baseline Average	Days Postirradiation 8	15	22	30	37	44	50-56
Control								
C49	4.8	--	--	--	--	--	--	--
C54	5.8	--	--	--	--	--	--	--
D87	5.7	3.6	--	--	--	--	--	--
10H	5.2	3.8	--	--	--	--	--	--
E10	5.7	3.6	--	--	--	--	--	--
E69	4.8	3.9	--	--	--	--	--	--
Chemicals								
B20	4.2	--	--	--	--	--	--	--
B52	5.1	--	--	--	--	--	--	--
C83	5.4	4.6	--	--	--	--	--	--
E12	4.9	4.5	--	--	--	--	--	--
E00	3.9	4.3	4.0	5.2	5.2	4.1	3.3	4.8
C61	5.4	4.5	4.6	5.0	5.2	4.7	4.5	4.6
BM II								
107E	N. R.	--	--	--	--	--	--	--
D63	4.9	--	--	--	--	--	--	--
B54	N. R.	5.8	--	--	--	--	--	--
20K	5.1	4.6	--	--	--	--	--	--
D61	5.2	3.7	3.4	3.2	--	--	--	--
49H	4.0	3.2	4.5	3.3	4.0	3.8	2.8	3.4
Ch + BM I								
B40	3.7	--	--	--	--	--	--	--
B46	3.7	5.7	--	--	--	--	--	--
B50	4.2	5.0	5.2	--	--	--	--	--
B18	5.4	3.6	4.5	4.3	4.0	3.4	5.1	3.6
B32	4.8	4.6	5.8	5.0	4.7	4.2	4.1	5.2
C45	5.4	5.2	5.2	4.7	3.4	4.2	5.2	4.9
Ch + BM II								
B59	3.4	4.3	3.0	3.8	4.5			
B60	4.6	3.6	2.8	3.5	4.2			
C24	3.8	3.8	3.6	4.0	4.5			
C34	4.1	3.2	4.6	5.5	7.5			
C59	3.6	4.3	3.6	5.3	4.2			
D50	4.6	3.2	4.2	4.5	5.5			

TABLE X

Sodium as Milliequivalents per Liter of Serum

Treatment and Animal #	Baseline Average	Days Postirradiation 8	15	22	30	37	44	50-56
Control								
C49	150.4	--	--	--	--	--	--	--
C54	156.6	--	--	--	--	--	--	--
D87	152.2	145.0	--	--	--	--	--	--
10H	143.2	155.0	--	--	--	--	--	--
E10	160.5	146.0	--	--	--	--	--	--
E69	151.6	148.0	--	--	--	--	--	--
Chemicals								
B20	153.5	--	--	--	--	--	--	--
B52	155.0	--	--	--	--	--	--	--
C83	151.0	147.0	--	--	--	--	--	--
E12	152.0	147.0	--	--	--	--	--	--
E00	148.0	149.0	153.0	145.0	144.0	156.0	149.8	163.3
C61	154.4	151.5	151.0	153.0	143.0	158.0	156.0	161.0
BM II								
107E	N. R.	--	--	--	--	--	--	--
D63	147.2	--	--	--	--	--	--	--
B54	N. R.	153.0	--	--	--	--	--	--
20K	149.4	143.0	--	--	--	--	--	--
D61	145.6	141.0	151.0	153.0	--	--	--	--
49H	155.4	143.0	146.0	155.0	151.0	168.0	163.0	158.0
Ch + BM I								
B40	150.0	--	--	--	--	--	--	--
B46	156.0	150.0	--	--	--	--	--	--
B50	149.0	144.0	144.0	--	--	--	--	--
B18	149.0	123.7	147.6	144.0	156.0	146.0	159.0	151.8
B32	156.0	150.0	150.0	144.0	167.5	160.0	163.2	154.0
C45	155.1	145.0	152.0	152.0	151.5	157.0	158.6	160.0
Ch + BM II								
B59	157.8	154.0	161.0	163.0	200 +			
B60	156.5	149.0	160.0	156.0	168.0			
C24	151.0	152.0	160.0	161.0	171.0			
C34	155.6	160.0	163.0	200 +	200 +			
C59	154.5	157.0	160.0	194.0	168.0			
D50	154.6	154.0	153.0	163.0	176.0			

TABLE XI

Chloride as Milliequivalents per Liter of Serum

Treatment and Animal #	Baseline Average	Days Postirradiation 8	15	22	30	37	44	50-56
Control								
C49	120.6	--	--	--	--	--	--	--
C54	117.6	--	--	--	--	--	--	--
D87	113.0	115.1	--	--	--	--	--	--
10H	116.8	115.1	--	--	--	--	--	--
E10	116.6	109.2	--	--	--	--	--	--
E69	121.8	111.8	--	--	--	--	--	--
Chemicals								
B20	115.1	--	--	--	--	--	--	--
B52	113.3	--	--	--	--	--	--	--
C83	119.4	111.8	--	--	--	--	--	--
E12	119.0	105.9	--	--	--	--	--	--
E00	122.6	118.5	126.7	112.6	116.8	117.4	116.5	113.3
C61	125.0	110.1	122.5	121.0	110.1	122.3	122.3	111.7
BM II								
107E	N. R.	--	--	--	--	--	--	--
D63	113.4	--	--	--	--	--	--	--
B54	N. R.	121.0	--	--	--	--	--	--
20K	122.3	105.8	--	--	--	--	--	--
D61	107.6	115.9	109.2	94.1	--	--	--	--
49H	123.5	101.7	118.3	108.4	114.3	109.9	122.3	109.2
Ch + BM I								
B40	119.2	--	--	--	--	--	--	--
B46	115.8	120.8	--	--	--	--	--	--
B50	107.5	110.8	102.5	--	--	--	--	--
B18	114.1	94.2	116.8	115.1	115.7	106.6	115.8	111.2
B32	121.8	114.2	116.0	63.9	120.7	117.4	117.5	118.8
C45	119.5	98.3	113.3	116.8	112.4	104.1	119.2	117.1
Ch + BM II								
B59	111.6	117.2	109.1	106.3	120.5			
B60	116.8	112.9	112.4	107.6	129.0			
C24	113.3	113.8	115.7	114.8	122.2			
C34	112.0	114.6	114.9	112.7	101.7			
C59	114.6	112.9	121.5	110.6	117.9			
D50	118.4	114.6	111.6	112.1	111.9			

FIGURE LEGENDS

Figure 1. Bone marrow being extracted from the femur of a monkey.

Figure 2. The Apparatus for the extraction of marrow from the long bones of a rhesus primate.

Figure 3. Survival of treated-irradiated Macaca mulatta at 30 days post-irradiation and at 360 days postirradiation:
Ch = Chemicals only; BM I = Bone Marrow only, I;
BM II = Bone Marrow only, II; Ch + BM I = Chemicals plus Bone Marrow, I; Ch + BM II = Chemicals plus Bone Marrow, II.

Figure 4. Hemogram showing the changes with time postirradiation of white blood cells, polymorphonuclear neutrophils, lymphocytes, and platelets.

Figure 5. Hemogram showing the changes with time postirradiation of red blood cells, reticulocytes, hemoglobin, and hematocrit.

Figure 6. Changes in the difference between water intake and urine water output with time postirradiation.

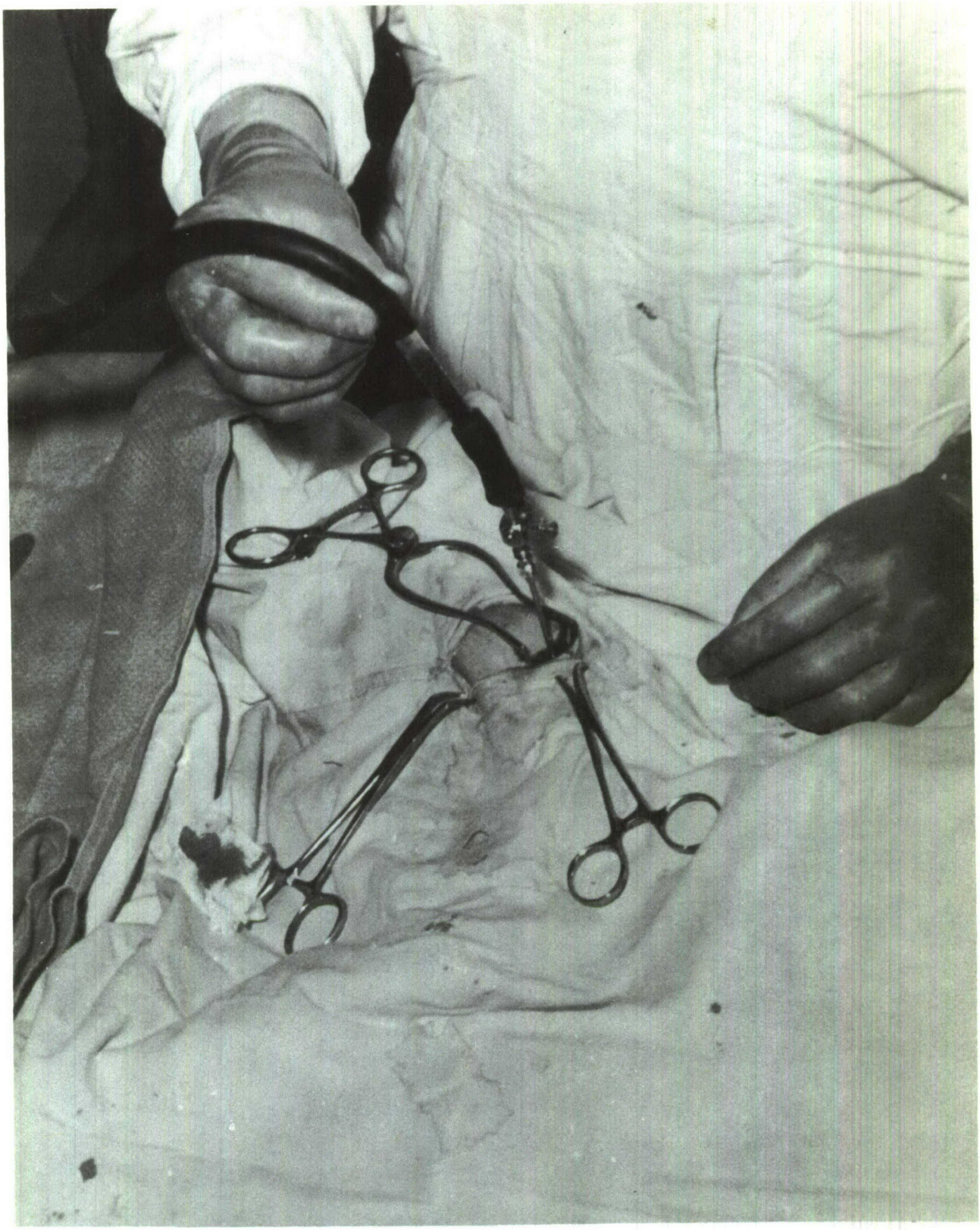


Figure 1

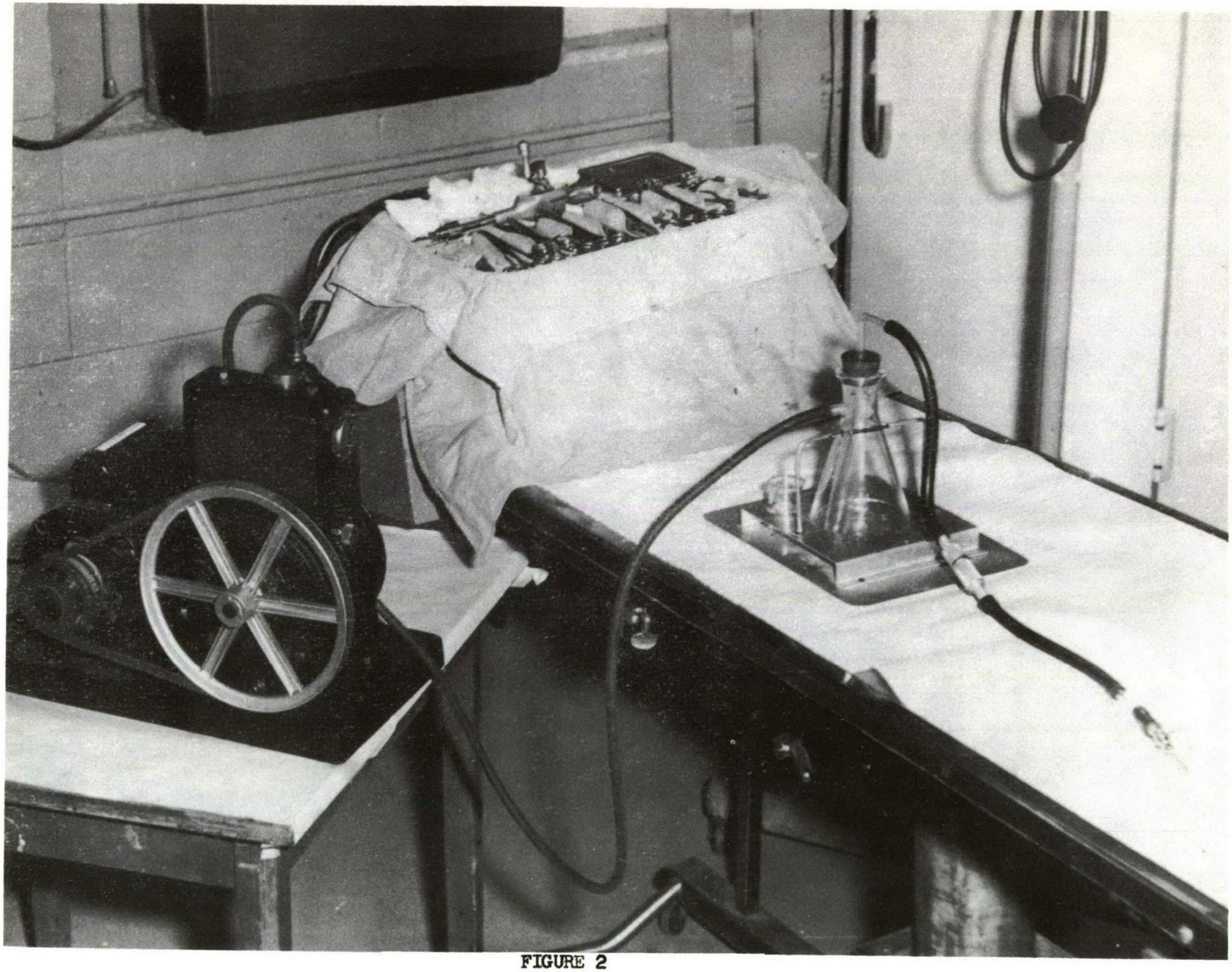


FIGURE 2

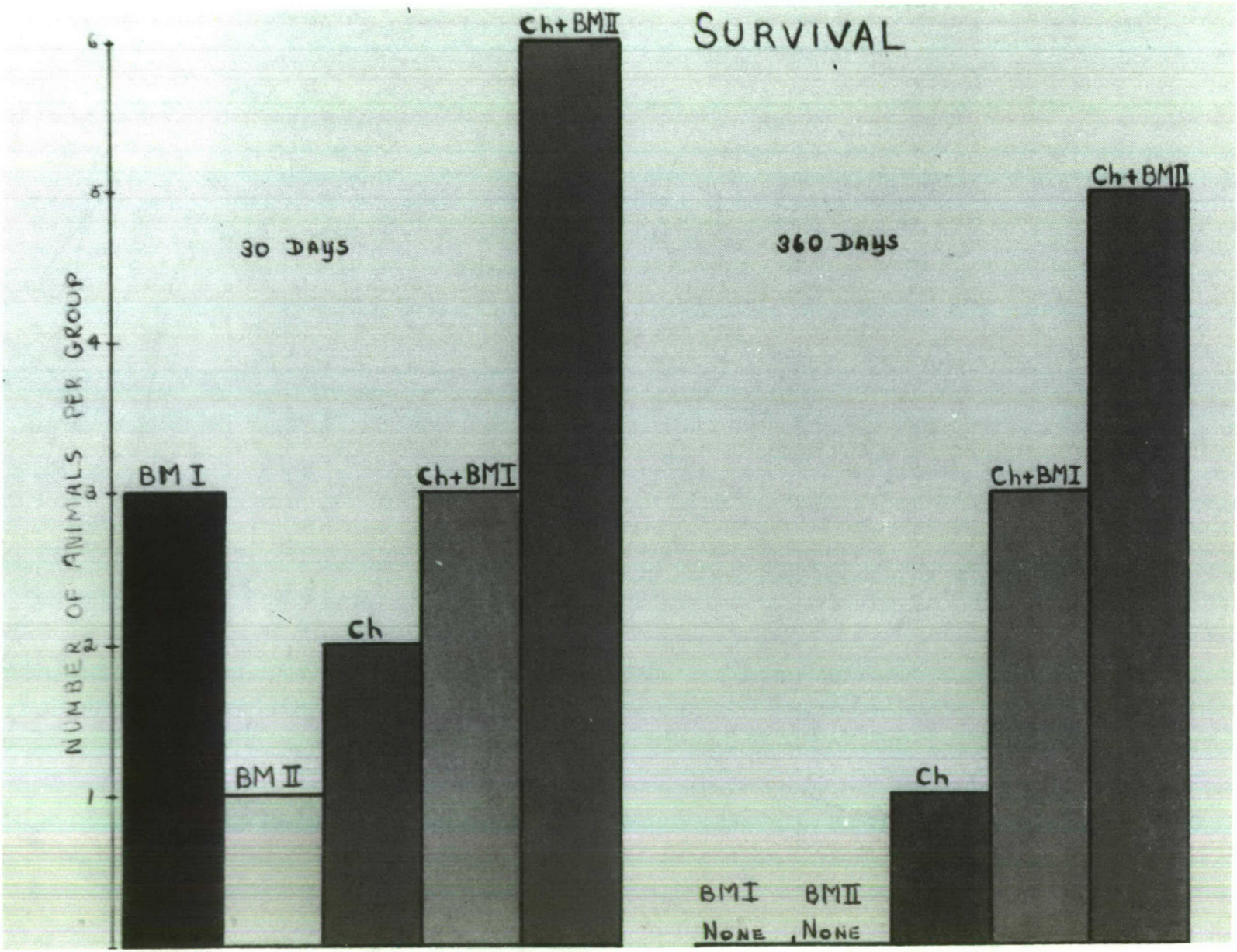

SURVIVAL
30 DAYS
360 DAYS
NUMBER OF ANIMALS PER GROUP
6
5
4
3
2
1
BM I
BM II
Ch
Ch+BMI
Ch+BMII
BMI
BMII
NONE
NONE

Figure 3

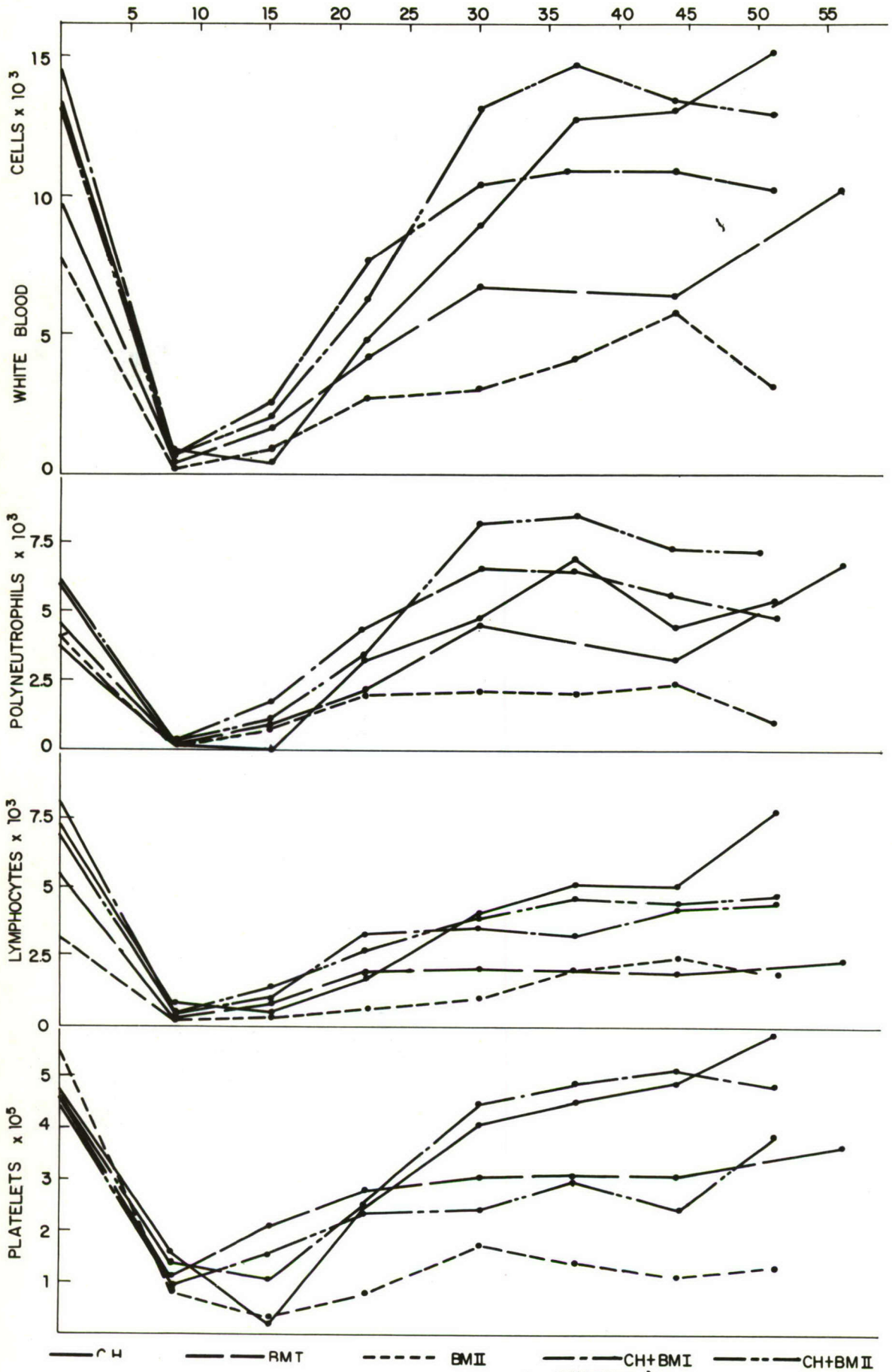

5
10
15
20
25
30
35
40
45
50
55
WHITE BLOOD CELLS x 10³
15
10
5
0
POLYNEUTROPHILS x 10³
7.5
5
2.5
0
LYMPHOCYTES x 10³
7.5
5
2.5
0
PLATELETS x 10⁵
5
4
3
2
1
CH
BMT
BMII
CH+BMI
CH+BMII

FIGURE 4

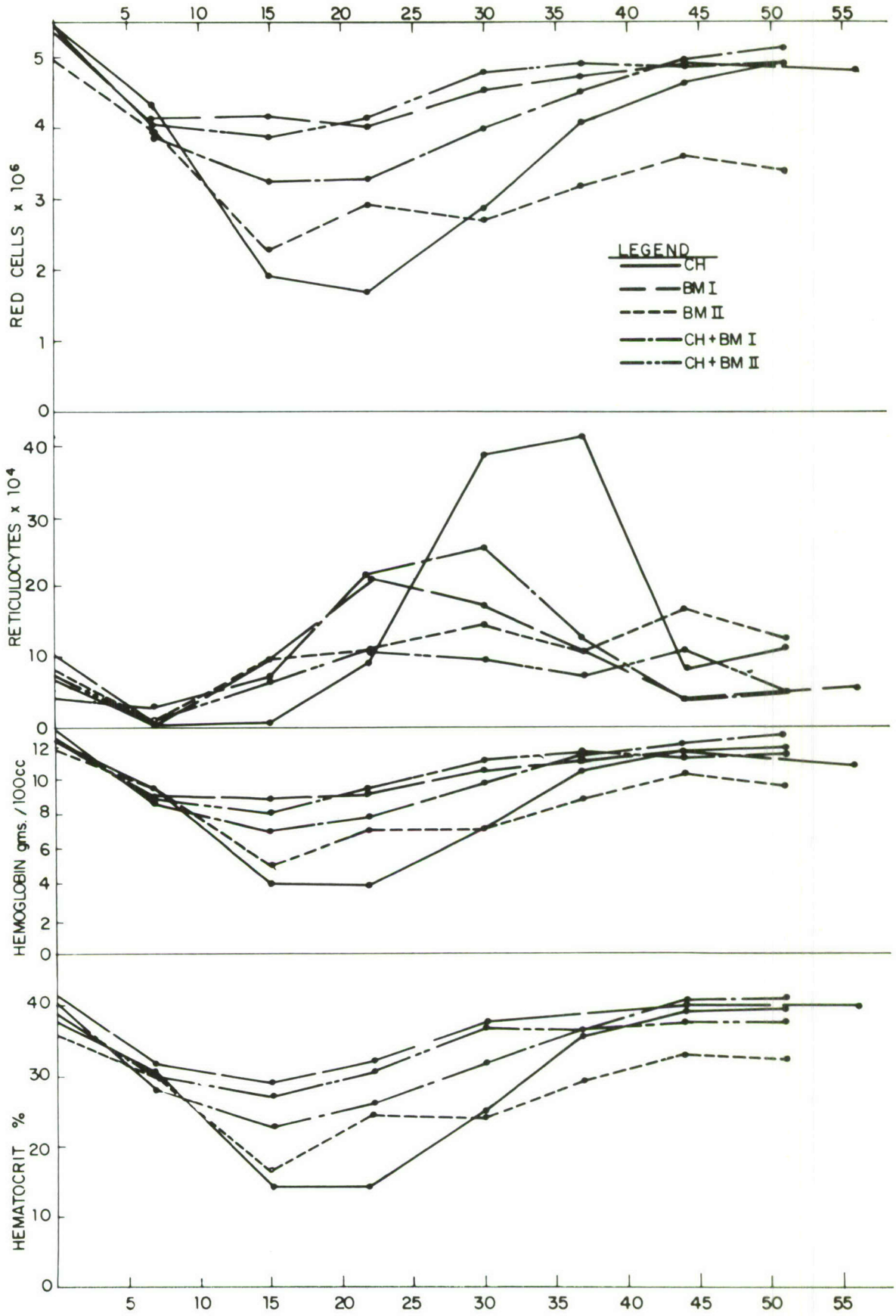

RED CELLS x 10⁶
RETICULOCYTES x 10⁴
HEMOGLOBIN gms. / 100cc
HEMATOCRIT %
LEGEND
CH
BM I
BM II
CH + BM I
CH + BM II

FIGURE 5

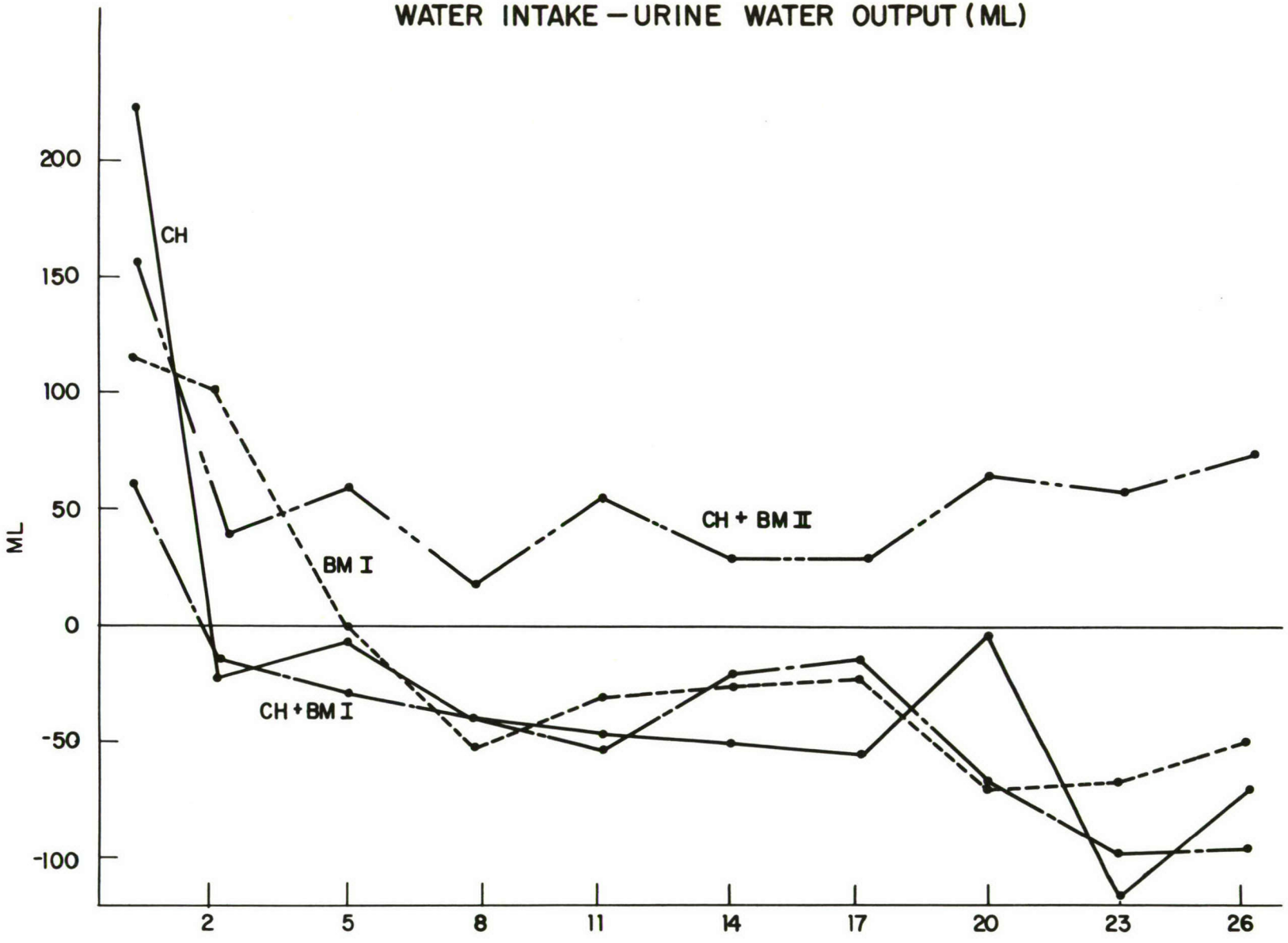

WATER INTAKE – URINE WATER OUTPUT (ML)
200
150
100
50
0
-50
-100
ML
CH
BM I
CH + BM II
CH + BM I
2
5
8
11
14
17
20
23
26
DAYS POSTIRRADIATION

FIGURE 6

11TH ANNUAL AIR FORCE SCIENCE & ENGINEERING SYMPOSIUM

(U) APPLICATION OF ENZYMES IN PHOTOCHROMIC OPTICAL SHUTTERS FOR NUCLEAR FLASH PROTECTION

SIDNEY ALLINIKOV
GS-13
AIR FORCE MATERIALS LABORATORY
(MAAE)
RESEARCH & TECHNOLOGY DIVISION
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

FOREWORD

This report was prepared by the Materials Engineering Branch of the Materials Applications Division. This work was initiated under Project Nr. 7381, Task Nr. 738101. It was administered under the direction of the Air Force Materials Laboratory, Research and Technology Division, Wright-Patterson AFB, Ohio, with Mr. Sidney Allinikov acting as project engineer.

TABLE OF CONTENTS

	PAGE
INTRODUCTION	1
LITERATURE REVIEW	2
FLASH BLINDNESS REQUIREMENTS AND PERFORMANCE OF DYE-ENZYME PHOTOTROPIC SYSTEM	8
PAPAIN-DYE INTERACTIONS	11
EFFECT OF PAPAIN ON PHOTOCHROMIC PERFORMANCE	13
Cyclic Fatigue	13
Color Decay Rate	16
Optical Density	18
DYE-ENZYME SYSTEM IN GELS AND FILMS	22
METHYLENE BLUE AND SPIROPYRANS	23
CONCLUDING REMARKS	24
REFERENCES	25
APPENDIX I	27
Experimental Procedures, Materials and Equipment	27

ILLUSTRATIONS

FIGURE		PAGE
1.	Triplet State Compound Energy Levels	6
2.	Optical Density Versus Transmittance	10
3.	Optical Density Versus Number of Flashes	14
4.	Decrease in Ratio of Time to Fade With and Without Papain	15
5.	Dye-Enzyme Solution Optical Density Versus Increased Methanol/Water Ratios	19
6.	Malachite Green-Glucose Oxidase Color Decay Rate with Increased Enzyme Concentrations	21
7.	Flash Photolysis Unit	28

TABLES

TABLE		PAGE
1.	Flash Blindness Protection Requirements and Dye-Enzyme Performance	9
2.	Fading Time (t_f) for Dye-Enzyme Solutions	17

INTRODUCTION

Intense electromagnetic radiation in the visible and ultraviolet spectrum emanating from thermonuclear explosions has introduced the problem of preventing eye damage to personnel in the area of such blasts. The normal blinking reflex time (150 milliseconds) is too slow to protect against flash blindness or choreoretinal burn when personnel are exposed to the destructive quantities of high intensity flashes of light generated at the time of the explosion. It is not difficult to imagine the perilous position of a pilot in a low flying high speed aircraft, were he suffering from flashblindness even for a period as short as 30 seconds. For this and other equally critical reasons, the Air Force and other agencies have been investigating different methods of providing quick response (microseconds) flash protection devices.

Photochromic materials with the ability to reach high optical densities under ultraviolet exitation and fade rapidly upon removal of the stimulating light source, have been a prime area of investigation for this purpose. This concept utilizes an ultraviolet flash source built into a visor or goggles that contain the photochromic material. If a nuclear blast occurs, a photoelectric cell triggers the UV flash, which in turn causes the visor or goggles worn by the air crew member to turn opaque. The relatively small effort spent in evaluating and adapting dye-enzyme systems for this application has already yielded good results. There is reason to believe other applications and other useful dye-enzyme systems will be revealed as this work is continued.

LITERATURE REVIEW

The terms "phototropy" and "photochromic" have been applied to the phenomenon of a reversible change of color induced by colored lights, as, in fulgides and the change in or loss of color of dyestuffs in light of specific wavelength. The term has been used by chemists, botantists, biologists, and biochemists for many years to explain this phenomenon of color change of solids and solutions. Brown (Reference 1) reviews the articles in Chemisches Zentrablatt from 1880 to 1910 and in Chemical Abstracts from 1907 to 1955 which deal with the phototropic phenomenon.

Most of the phototropic substances are organic in nature, although many inorganic compounds have been found to exhibit phototropy. Several hundred compounds are known to be phototropic and most of these exhibit the phenomenon in the solid state. Among these solid compounds are anils or Schiff bases, disulfides, fulgides, hydrazones, spiranes, osazones, semicarbazones, stilbene derivatives, and other numerous organic compounds.

Several reviews in different areas of phototropy have been published. Phototropy as related to chemistry up to about 1921 was reviewed by Stobbe (Reference 2); Chalkley (Reference 3) reviewed the field as related to chemistry up to about 1928: Bhatnagar, Kapur, and Hashmi (Reference 4) up to about 1938: Brown (Reference 2) up to about 1955. Van Overbeek (Reference 5) reviewed the biological field up to about 1938. No information appears to be available on the relation of enzymes in the phototropic process.

Diemar and others (References 6-10) give a comprehensive review in their articles on artificial colors and enzyme reactions. They have studied in great detail the effects of artificial dyes upon catalase, pepsin, and stomach juices, trypsin and the intestinal juices. Their conclusions are important with respect to the artificial coloring of foods.

Braun (Reference 11) has studied the interaction of protein with triphenylmethane dyes. The triphenylmethane dyes were found to be rendered colorless by myosin and nucleoproteins. When denatured by heat, urea, or ultraviolet rays, the proteins were found to possess still greater decolorizing action which was attributed to the combining of the dyes with the proteins, the points of attachment being the sulfhydryl (-SH) groups. This reaction could be inhibited by the action of mercury salts or oxidation of the -SH groups by hydrogen peroxide. Through spectrophotometric studies of the decolorization of triphenylmethane dyes with proteins in alkaline medium, Braun (Reference 11) showed that, during decolorization of triphenylmethane dyes with proteins, the amino acid groups combine with the dye.

A number of investigators have attempted to interpret the cause of phototropy in solutions (References 12-21). The dyes most commonly used to explain the mechanism of this phenomenon have been the leucocyanides. The phototropy of malachite green leucocyanide has been investigated in dry alcohol and also in mixtures of alcohol and water by a number of authors (References 22-28). The results show that on irradiation, malachite green leucocyanide dissolved in alcohol is photo-ionized to

yield the carbonium $(MG)^+$ and the cyanide ion, and that the quantum efficiency of the process is very close to unity. Removal of the light source causes the color of the solution to fade.

Probably the best explanation of the mechanism involved in the process is given by Holmes (Reference 29) who examined the absorption spectra of the products of dark reaction formed under various conditions and compared them with those of the leucocyanide, the carbinol, and the ethyl ether. The proposed mechanism which accounts for the formation of the various dark reaction products according to the principles of mass action are given below:

$$MGOH \underset{r_4}{\overset{r_1}{\rightleftharpoons}} (MG)^+ + OH^-$$

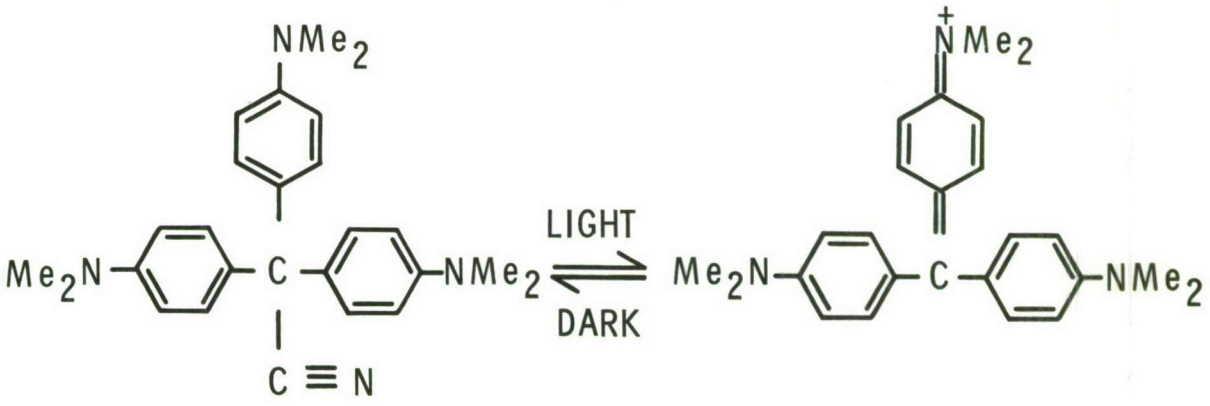
$$MGCN \underset{r_5}{\overset{r_2}{\rightleftharpoons}} (MG)^+ + CN^- \quad \begin{matrix} HCN \\ + \\ H_2O \\ \uparrow\downarrow \\ + \\ C_3H_5OH \end{matrix}$$

$$MGOC_2H_5 \underset{r_6}{\overset{r_3}{\rightleftharpoons}} (MG)^+ + OC_2H_5^- \quad \begin{matrix} \uparrow\downarrow \\ HCN \\ + \end{matrix}$$

Reactions r_1, r_2, and r_3 are all caused by the absorption of photons at wavelengths in the ultraviolet or nearly-visible region whereas reactions r_4, r_5, r_6 are all spontaneous reversible thermal reactions.

A review of literature indicates most of the effort on development of phototropic systems has involved either bond breaking mechanisms, or excitation of a molecule to a metastable electronic state (Reference 30-31). In the former category, absorption of light causes breaking of a chemical bond. This basically produces a compound with a different absorption spectrum than the previous structure. The triphenylmethane leuco-dyes and spiropyrans are examples of this class of compounds which have received wide attention.

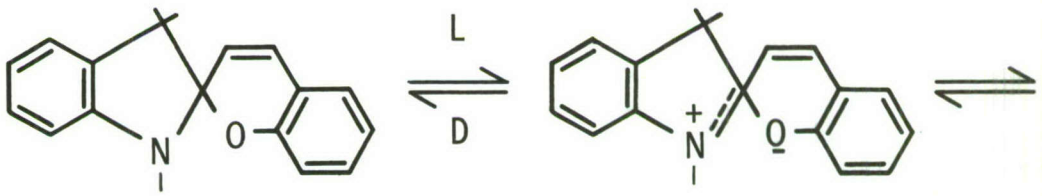
PHOTOTROPIC TRIPHENYLMETHANE SYSTEM (LEUCONITRILE)



COLORLESS LEUCONITRILE

(COLORED ION)

INDOLINOSPIRORYRANS



MESOMERIC & ISOMERIC FORMS

In other class of compounds such as pentracene, naphthalene, and dibenzanthracene, where color change is due to excitation of the molecule, chemical bonds are not broken. Absorption of the light raises the molecule to an excited singlet state, from which it drops, by a non-radiative pathway to metastable level, with an accompanying change in absorption spectrum from the original form.

The energy level diagram for a triplet state compound is shown in Figure 1.

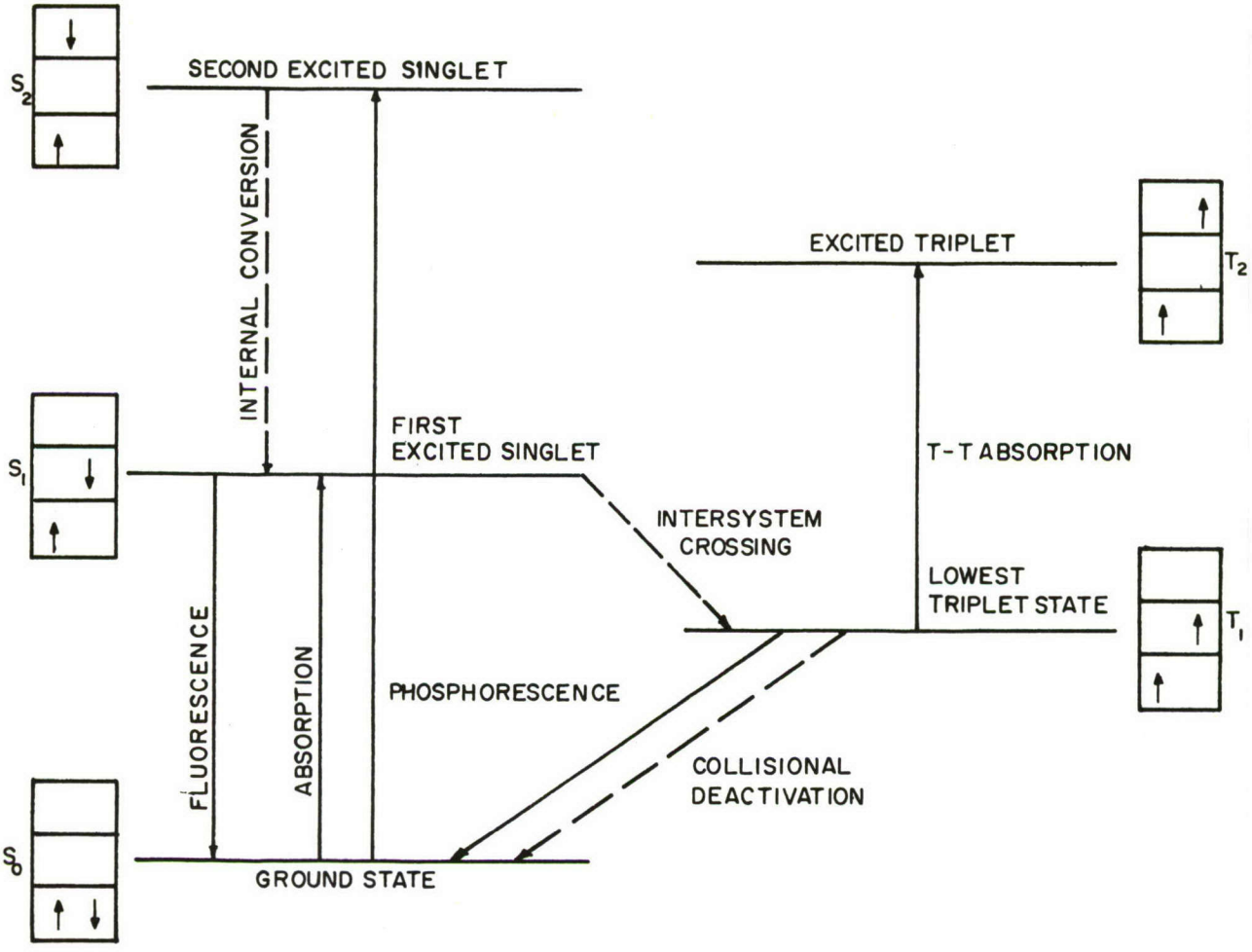


Figure 1. Triplet State Compound Energy Levels (Ref 1)

The molecules are raised to the first excited singlet state from the ground state by ultraviolet light absorption. These in turn, pass to the lowest triplet state. Visible light is absorbed by the triplet state form of the molecule. Decay of the triplet molecule can also occur through phosphorescence and by collision with other molecules. Arrows in boxes indicate electron spin.

Most phototropic systems being investigated suffer disadvantages which must be overcome if they are to be of use as an automatically responsive filter. The spiropyrans appear to have low quantum efficiencies. They are temperature sensitive, and exhibit fatigue on repeated cycling. The aromatic hydrocarbons show high promise for future development. However, little data are presently available on quantum efficiencies, extinction coefficients, and absorption spectrums of this class of compounds. They are also reported to be somewhat oxygen sensitive, losing phototropic properties after several months exposure.

The phototropic triphenylmethane dyes are capable of reactions with quantum yields approaching unity in some solvents. Very high extinction coefficients, and broad spectral coverage are additional properties that are favorable for the flash type application. Disadvantages of these dyes, however, have been slow reversal time from colored to colorless state, fatigue during cyclic exposure, and difficulty of incorporation into plastic or glass media. The detailed discussions of results shows how the enzyme dye combination alleviates cyclic fatigue and decreases reversal time.

FLASH BLINDNESS REQUIREMENTS AND PERFORMANCE OF DYE-ENZYME PHOTOTROPIC SYSTEM

A photochromic compound, to effectively protect against flash blindness, must be capable of converting from a transparent state to a high optical density very rapidly. The compound would be required to remain opaque in the presence of harmful radiation, and then rapidly revert to the colorless state upon attenuation of the harmful radiation to a safe level. The requirements for the desired flash protection system and relative performance of the dye-enzyme compound developed to date are shown in Table I.

TABLE I

FLASH BLINDNESS PROTECTION REQUIREMENTS AND DYE-ENZYME PERFORMANCE

Condition	Requirements	Performance
Open State transmittance, visible	80%	45%
Open state transmittance, UV	5%	--*
Open state transmittance, IR	10%	Side Band Filter Acceptable
Haze	5%	--*
Closed state transmittance, visible	.01%	.01% - .05%
Maximum density, UV	3.0	Side Band Filter Acceptable
Maximum density, IR	3.0	Side Band Filter Acceptable
Response time (to secure max density)	75 microseconds	10 microseconds
Response time (to secure max transmittance)	25 sec	Milliseconds to minutes
Operating temperature range	17°C-57°C	
Fatigue performance	50 cycles	75 cycles
Open state transmittance, visible, in direct sunlight (minimum)	15%	--*

* Unknown

Optical density, as used in this report, refers to the degree of opaqueness to be achieved by the compounds during various stages of operation of the shutters. The degree of opaqueness, or optical density, would be very low for normal flight conditions, and as high as possible during periods of high intensity flashes of light. The curve below shows the relationship between optical density and light transmittance.

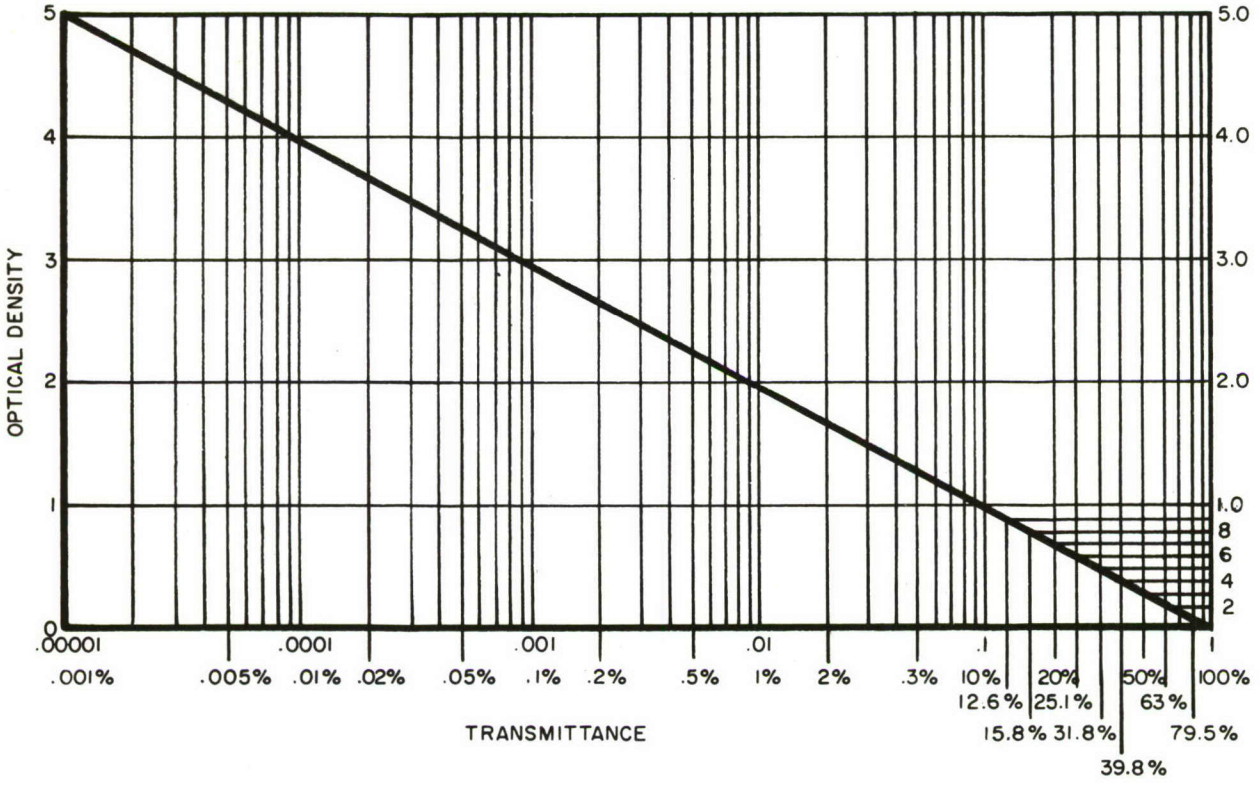


Figure 2 Optical Density Versus Transmittance

PAPAIN-DYE INTERACTIONS

Enzymes are proteinaceous materials consisting of high molecular weight copolymers of -amino acids and imino acids linked by amide bonds. They have a highly ordered structure consisting of primary, secondary, and tertiary structures. The sequence of amino acids in the polymer chain comprises the primary structure. The secondary structure consists of a -helix formation by hydrogen bonding between the carbonyl and amino groups of two amino acids. A tertiary structure is formed in the way the polymer chain is folded.

The ability of enzymes to act as very effective catalysts is well known. Although many enzymes are specific in regard to substrates, papain has been used with a wide variety of natural and synthetic substrates successfully. Because of these characteristics, more data is available on the properties of papain than many other enzymes. Formation of a complex between the enzyme and its substrate always occurs as a preliminary step in an enzymatic reaction. Formation of the complex is attributed to the forces of attraction between polar and apolar groups on the substrate and amino acids in the enzyme. The complex formed with substrate and enzyme insures the proper spatial relationship for reaction to occur. That portion of enzyme which provides catalytic activity is known as the active site. In the case of papain, sulfhydryl group is referred to as the active of that particular enzyme. Amino acids at the active site participate in making or breaking chemical bonds in the substrate with subsequent formation of the reaction product. The process can be repeated again upon release of the products of reaction from the active site.

The interactions between dyes and enzymes have been demonstrated by several investigators, as indicated in the literature review contained in this report. In view of this data, and observations of results, it seems likely that an enzyme-dye complex is being formed as a first step. This action would be expected in following the usual enzyme mechanism where it acts as a catalyst.

One or more additional steps in the process would then take place upon irradiation until the dye cation were produced. This action can be represented by the following scheme:

$$\underset{\text{(colorless)}}{[\text{Dye-Enzyme}]} \xleftrightarrow[\text{dark}]{\text{light}} \underset{\text{(colored ion)}}{\text{Dye}^{+}\quad \text{Enzyme}^{-}}$$

The presence of radiant energy in the process may of course alter the role of the enzyme as a catalyst. It may act as both an acceptor and a catalyst, although it should be noted that peak absorption wavelength of papain lies in the UV range beyond the 3000Å to 4000Å peaks intensity light sources used in the experiments. If the enzyme were acting in some capacity other than as a catalyst, it might be retaining the cation and anion after photodisassociation in close proximity to each other, thus acting as a cage. Needless to say, these postulations are greatly simplified, and may not represent the true mechanisms of the process.

EFFECT OF PAPAIN ON PHOTOCHROMIC PERFORMANCE

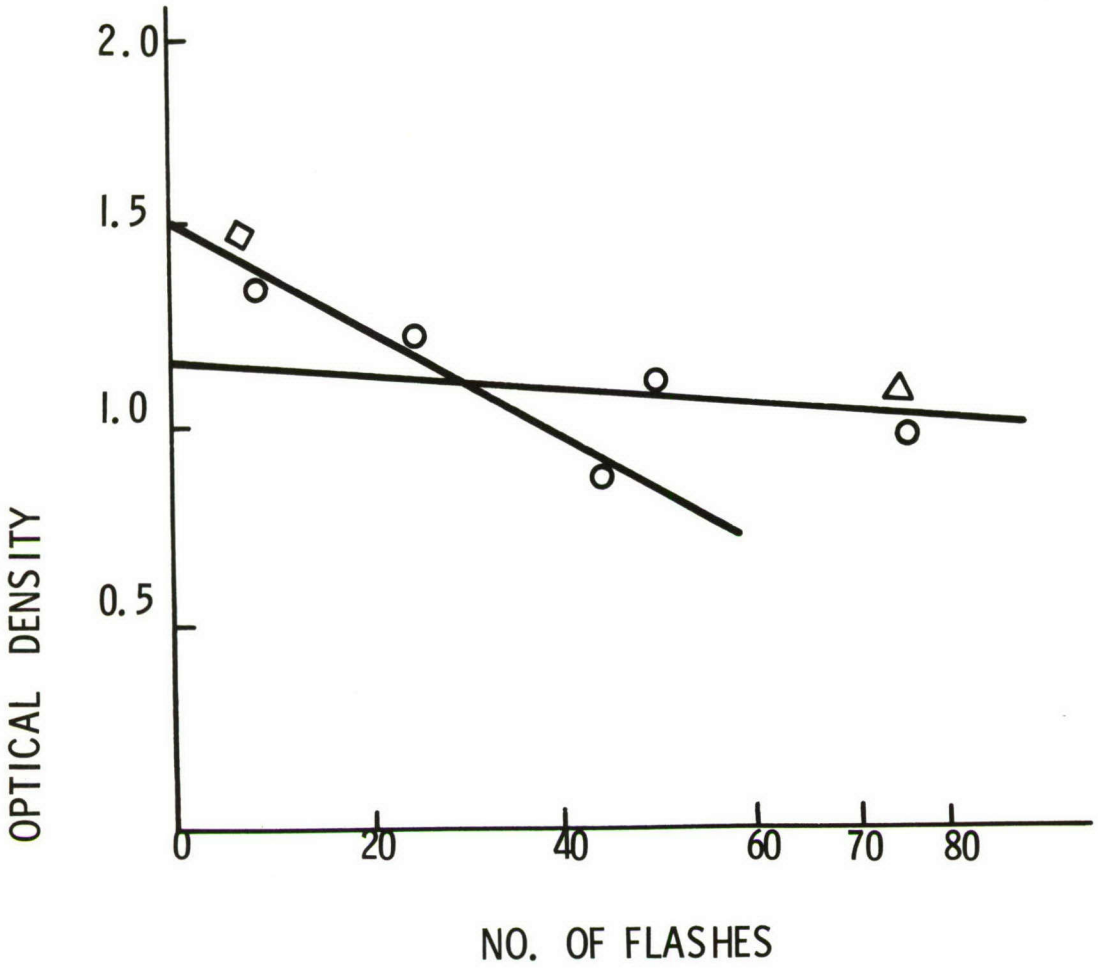
Cyclic Fatigue

A satisfactory photochromic compound must be able, upon repeated cyclic exposure to a light source, to reach the same optical density each time it is exposed. The compound should not decrease in rate of recovery to the leuco form, nor become unuseable optically because of turbidity. Evidence of failure in any of these factors can be classified as cyclic fatigue.

The presence of papain in combination with the leuco form of triphenylmethane dye in solution has a marked effect on improvement of cyclic fatigue of the compound. Quantitative experiments were performed on malachite green solutions with and without papain. "Concentrations" of compounds used were 1×10^{-4}N dye, and 1×10^{-3} sodium bisulfite with a bisulfite/dye ratio of 10. Papain "concentration" when present was 2.5×10^{-4}M. Each solution was flashed 70 times, and inspected for evidence of fatigue as defined above after each flash.

Figure 3 shows the change in optical density originally obtained with repeated flashes both with and without papain. The optical density initially obtained of a solution without papain at a bisulfite/dye ratio of 10 decreases more rapidly with each flash than does the optical density of the solution with papain. The ratio of $\frac{tf\ (+P)}{tf\ (-P)}$ also decreases with the number of flashes in solutions without papain, as shown in Figure 4.

OPTICAL DENSITY VS. NUMBER OF FLASHES

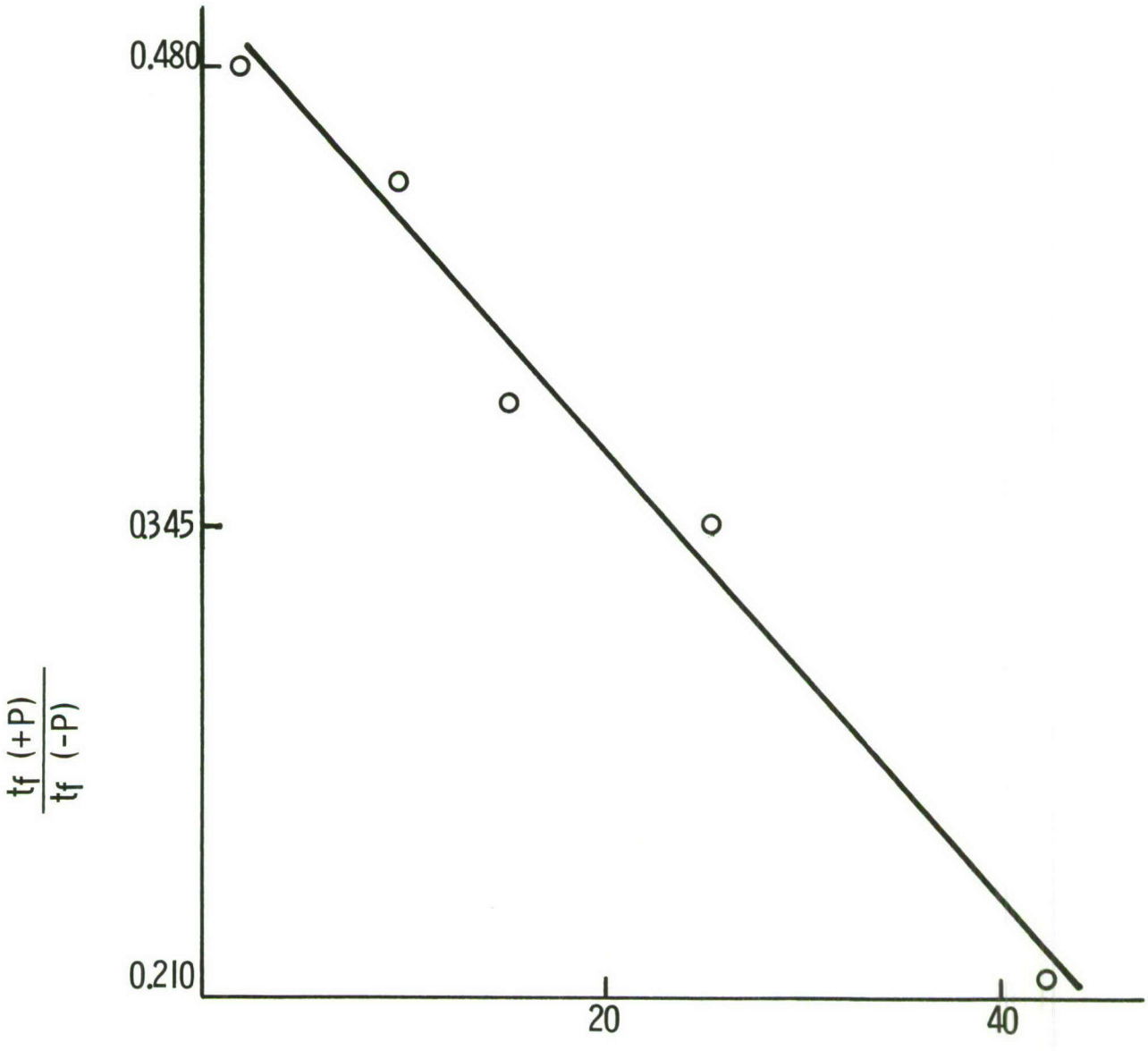


□ WITHOUT PAPAIN

△ WITH PAPAIN

Figure 3

DECREASE IN RATIO OF TIME TO FADE
WITH AND WITHOUT PAPAIN



t_f (+P) Time to fade with papain

t_f (-P) Time to fade without papain

Figure 4

COLOR DECAY RATE

The rate of color or optical density decay (time to fade from the colored to the colorless form) is an important factor in consideration of a photochromic system for flash blindness protection. Dye systems which might be acceptable otherwise have been notoriously poor performers in this respect. Some phototropic dyes are temperature sensitive, requiring application of heat to cause rapid reversion from the colored to leuco form. Other lose color very rapidly, but may phosphoresce for some time afterwards in the process. This phosphorescence has a decidedly adverse effect on vision. Some of the most responsive compounds with high quantum efficiencies are irreversibly opaque upon exposure to a flash source. It was found the decay rate was dependent upon several interrelated factors such as reducing agent concentration and pH as well as the presence of the enzyme. Table 2 presents experimental results where the pH and bisulfite/dye ratio were varied. Malachite green and papain concentration were held constant.

	pH							
(HSO_3^-)	7		6		5		4.5	
(Dye)	+P	-P	+P	-P	+P	-P	+P	-P
2	23[a]	V.L.	-	-	-	-	V.L.[b]	V.L.
5	2	V.L.	12	-	35	-	V.L.	-
12	2.8	30	7	-	-	-	V.L.	-
20	-	-	4.3	V.L.	-	-	-	-
42	-	-	-	-	25	V.L.	-	-
50	0.4	9.5	-	-	-	-	-	-
100	0.3	2	1.8	-	10	-	16	-
200	0.08	0.160	0.9	5	3.6	-	15	-
500	0.02	-	0.26	-	1.4	5	5.4	-
1000	0.006	-	0.13	-	0.8	-	2.6	-

a = t_f in seconds

b = V.L. (very long) (t_f few min)

TABLE 2. FADING TIME (t_f) FOR DYE-ENZYME SOLUTIONS

The conclusions drawn from the results in Table 2 are threefold. The presence of papain always shortens t_{fade} compared to the corresponding solution without papain. Increasing the bisulfite to dye ratio shortens t_{fade} at any pH. Time to fade is decreased with higher pH values for the same bisulfite/dye ratios.

OPTICAL DENSITY

The fundamental quality any photochromic compound must possess is the ability to achieve a high optical density in microseconds. The triphenylmethane dyes are excellent candidates in this respect. Four important advantages are possessed by this class of dyes:

1. Quantum efficiencies near unity in some solvents.
2. Very high extinction coefficients ($> 10^5$ molar decadic units).
3. Absorbs over broad spectral range.
4. Rise time of color formation is extremely rapid ($< 10\mu$ sec).

The role of papain in the light reaction (leuco to colored) has never really been determined because of the high speed of response of these dyes to light stimulus. In our experiments the rise time of color formation could not be distinguished from the rise time of the flash lamp, which was approximately 10 sec. Optical densities close to 4.0 were measured in a cm thick sample starting from an open state of 45% transmission. The flash apparatus used was a 200 watt/second source described in Appendix I. The extent of photochemical conversion of the dye compound under ultraviolet exitation was found to be highly dependent upon the solvent. The solvents used were ethanol-water, methanol-water, acetone-water, and benzene-methanol. Phototropy was present in all of the solvents tried, however, the methanol-water solution proved the most effective. The ratio of methanol to water as expected also had an influence on photochemical conversion of the dye form. Figure 5 presents the trend in optical density change with varying proportions of methanol and water. The samples were a 40 mil layer of dye-enzyme solution exposed to the 200/watt/second ultraviolet flash source, as described in Appendix I.

DYE-ENZYME SOLUTION OPTICAL DENSITY VS INCREASED MeOH/H_2O RATIOS

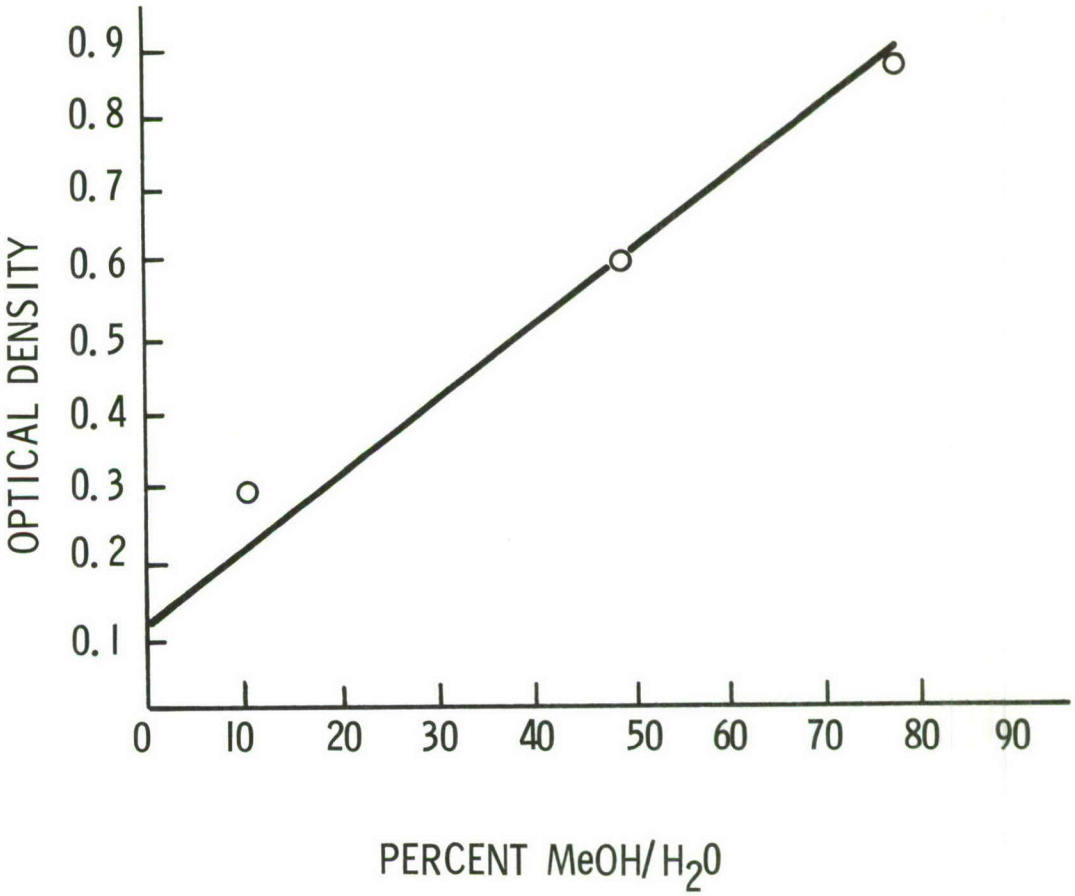


Figure 5

Several preliminary experiments were performed with urease and glucose oxidase in an attempt to note if there were similar phototropic effects with other enzymes containing SH groups. Urease had no effect although it contains a greater number of -SH groups per mole than papain. Urease, however, is classically specific in activity on urea, with no other substrate being known. The results with malachite green and glucose oxidase indicated phototropism and a markedly rapid back reaction rate. The time to revert from the colored to the open state decreased with increased concentration of glucose oxidase. Figure 6 presents this information.

MALACHITE GREEN-GLUCOSE OXIDASE
COLOR DECAY RATE WITH INCREASED
ENZYME CONCENTRATION

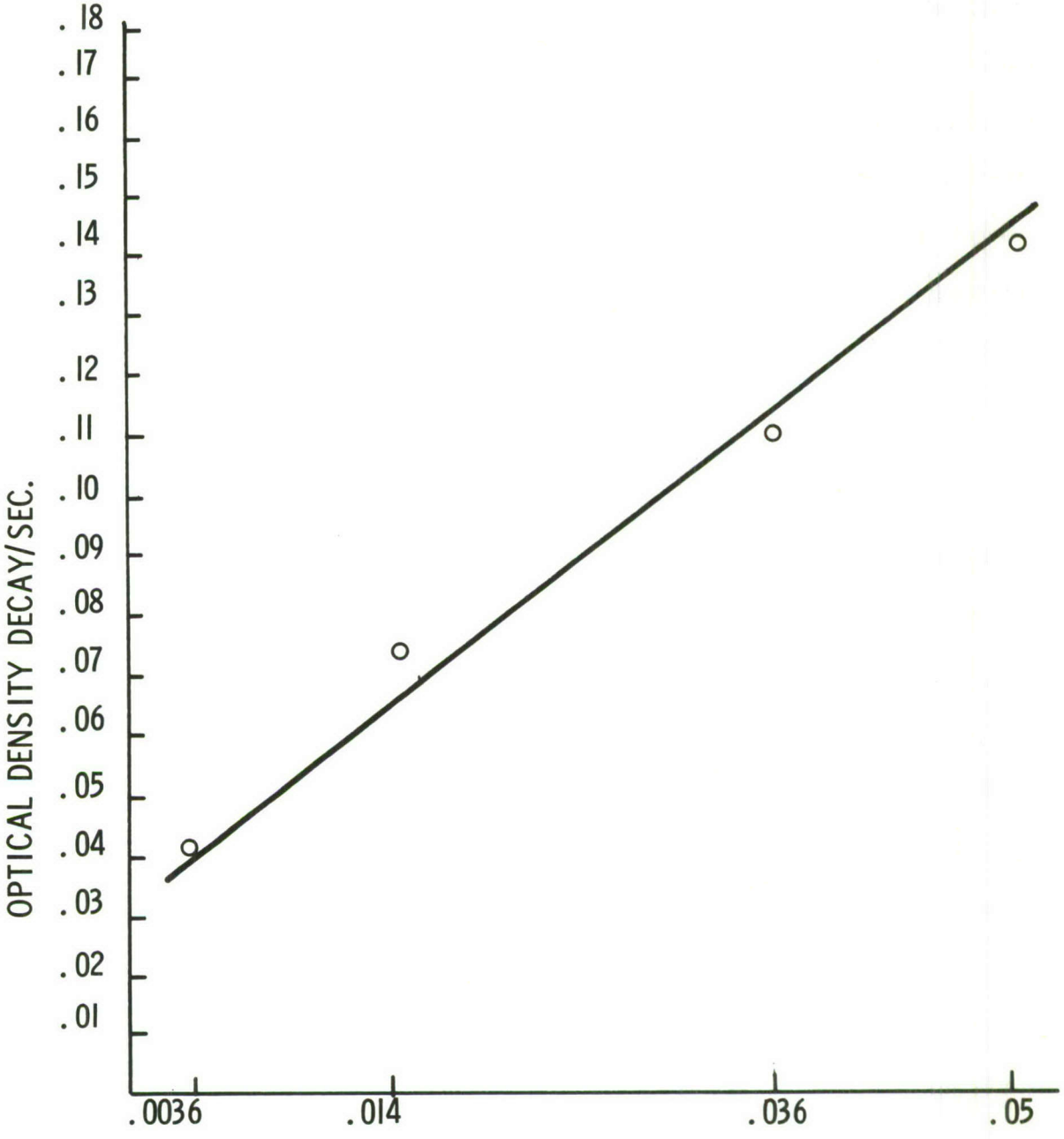


Figure 6

DYE-ENZYME SYSTEM IN GELS AND FILMS

Some preliminary work was done on attempting to incorporate the enzyme-dye system into polyvinyl alcohol and also in gel form. The purpose of this work was to obtain some idea if it were possible to incorporate the system into solid or semi-solid media. If this could be done successfully, it would function as either a coating or inter-layer for application to visors, glasses or windshields.

Reaction to ultraviolet was considerably slowed by the polyvinyl alcohol films, which was expected from results of previous investigators. The guar based gels however appeared to allow sufficient mobility for fast reaction rates. The guar gelling agent was also selected because of its inertness to the dye system. Another advantage of the gum used in making the gel was based on its self-complexing properties. A gel can be formed with these compounds by rapid agitation for several seconds at room temperature, thus precluding the necessity for boiling as done with many other gels. Denaturization of the protein was also avoided, by forming the gel at room temperature. As previously stated, this work was of a preliminary nature, and by no means exhausts the possibilities for incorporating the compound into optical media. Since a compound in coating form would be desirable some thought might be given to micro encapsulation of the system, thus retaining the advantages of a liquid formulation in solid form. Two typical formulations for a film and gel are given below.

Typical Formulations	Weight (gms)
1. Dye	0.03
Activated papain	1.6
Methanol	1.0
Polyvinyl alcohol	5.0
Water	100.0
2. Dye	0.03
Activated papain	1.6
Guar gum	1.0
Water	100.0

This phase of work was very preliminary and much remains to be done before the dye-enzyme system can be incorporated into solid media effectively.

METHYLENE BLUE AND SPIROPYRANS

Neither papain nor glucose oxidase proved to be effective with methylene blue or 8'-methoxy-6'-nitro 1,3,3-trimethyl-indoline-benzopyrylospiran. Methylene blue converted to the leuco form is an interesting dye because of its rapid color change capability under ultraviolet stimulus. This particular dye also reverts from the colored to leuco form when exposed to radiation in the visible spectrum. Neither enzyme had any effect at all on this dye. Glucose oxidase showed a definite inhibiting effect on the spiropyran phototropic characteristics. Quantities of the enzyme added to the dye in toluene radically increased time for color change and the reverse action.

CONCLUDING REMARKS

The experimental data discussed herein shows that a dye-enzyme photochromic solution can be made which has a good potential in solving a difficult Air Force Applications problem. Perhaps of equal importance is the fact a new "applications area", that of applying dye-enzyme compounds as functional engineering materials is presented.

In addition to the applications usually suggested for photochromic materials, such as computer memory storage units and sunglasses, other applications suggest themselves. Preliminary investigation indicates the sensitivity and specificity of these compounds may be used to detect minute quantities of toxic organic compounds or their decomposition products. Their presence of the compounds could be signaled by a color change.

Photochromic solutions are now being introduced as "shutters" for laser applications. Enzyme modified dyes could easily play a part in this application. Photoelectric properties of dye-enzyme systems worked with on a preliminary basis also show possibilities as detectors for various wavelengths of electromagnetic radiation. These applications are to be explored for feasibility in Air Force operational systems.

REFERENCES

1. Brown, Glenn H., WADC Tech. Report No. 59-436 (1959)

2. Stobbe, H., Ber. Verhandl. sachs Akad. Wiss. Leipzig. Math Phys Klasse, 74, 161-250 (1922)

3. Chalkley, L., Chem. Reviews, 6, 217-280 (1929)

4. Bhatnagar, Kapur and Hashmi, J. Indian Chem. Soc. 15, 573 (1938); Chem. Abstracts, 33, 3698 (1939)

5. Van Overbeek, J. Botan Reviews, 5, 655 (1939)

6. Diemar, W. and Hausser, H., Z. Leberism-Untersuch u-Forsch 92, 165 (1951)

7. Diemar, W. and Bockhoff, K., Z. anal. chem., 139, 25 (1953)

8. Diemar, W. and Hausser, H., Z. Leberism-Untersuch u-Forsch. 92, 403 (1951)

9. Diemar, W. and Z. anal. chem. (139). 35 (1953)

10. Diemar, W. and Bockhoff, K., ibid. 139, 267 (1953)

11. Braun, A.D., Biokhimiya. 16 399-409 (1951)

12. Braun, A.D., Savost'yanova, M.V. and Morozova, R.I., Izvest. Akad. Nauk USSR., Ser. Fiz. 14, 536-41 (1950)

13. Lifschitz, J., and Joffe, C.L., Ber 52B, 1919-1926 (1919)

14. Schlenk, W. and Herzenstein, A., Ber.43, 3541-3546 (1910)

15. Singh, B.K., J. Am-Chem-Soc., 43, 333-334 (1921)

16. Singh, B.K., and Bhaduri, B., Trans. Faraday Soc., 27 478 (1913)

17. Von Mechel, L. and Stauffer, H., Helv. Chim Acta 24, 151 (1941)

18. Brode, W.R. and Wyman, G.M., J. Research NaH. Bur Standards. 47. 170 (1951)

19. Brode, W.R., Gould, J.H., and Wyman, G.M., J. Am. Chem. Soc., 74, 4641 (1952)

20. Chaude, G. and Rumpf, P., Compt. rend. 236, 697-699 (1953)

21. Livingston, R. and Ryan, V.A., J. Am. Chem. Soc., 75, 2176-2181 (1953)

22. Lifschitz, J., Ber., 52, (1919)

23. Lifschitz, J. and Joffe, C.L., ibid, 97, 426 (1921)

24. Weyder, E. and Frankenburger, W., Trans. Faraday Soc., 27, 561 (1931)

25. Weyder, E., Frankenburger, W., and Zimmerman, W., Z. physik. Chem., 317, 276 (1932)

26. Harris, L. and Kaminsky, J., J. Am.Chem.Soc., 57, 1151, 1154 (1935)

27. Germann, F.E., Gibson, C.L., ibid, 62, 110 (1940)

28. Calvert, J.G. and Rechen, H.E., ibid, 74, 2101 (1952)

29. Holmes, E.O., ibid, 61, 434-441 (1957)

30. Study of the Triplet State in Fluid Solvents, Porter and Windsor, Disc. Faraday Soc., 1954, 17, 178

31. The Triplet State in Fluid Media, Porter and Windsor, Proc. Roy. Soc. 1958, 245A, 238

APPENDIX I

EXPERIMENTAL PROCEDURES, MATERIALS AND EQUIPMENT

Experiments on the effects of different bisulfite concentrations and pH conditions were performed by the Space Technology Laboratories, Redondo Beach, Calif. and Magna Products Inc., Anaheim, Calif. under Contract AF 33(657)-11708 to the Air Force Materials Laboratory for the purpose of optimizing the dye-enzyme photochromic system.

The enzymes papain, urease and glucose-oxidase were purchased from the California Corporation for Biochemical Research. The papain was activated with bisulfite prior to accomplishing an experiment. The other enzymes were used as received. The other compounds including dyes and solvents were reagent grade, drawn from Air Force supplies.

PROCEDURES AND EQUIPMENT

Enzyme-dye solutions were prepared by dissolving quantities of the dye in the solvent of choice and then the addition of the enzyme previously prepared in the same solvent or a cosolvent as required. The solutions were stirred for 10 minutes at room temperature and filtered, centrifuged, and filtered. Adjustments and measurements of pH were made with use of a Beckman "zeromatic" pH meter.

Steady state illumination was carried out with a standard General Electric 275 watt ultraviolet lamp. Optical density of solutions were measured before and after steady state illumination with a Bausch and Lomb photometer. Flash experiments including optical density and time measurements related to those experiments were made as described below.

The samples were exposed to a 200 watt/sec ultraviolet flash source. Peak intensity occurred in 25 microseconds, with a 50 percent fall off from peak intensity in 65 microseconds from zero time. The instrumentation used could detect and record the difference and time for change in optical density in the visible range. The block diagram presented in Figure 7 indicates the general instrumentation arrangement.

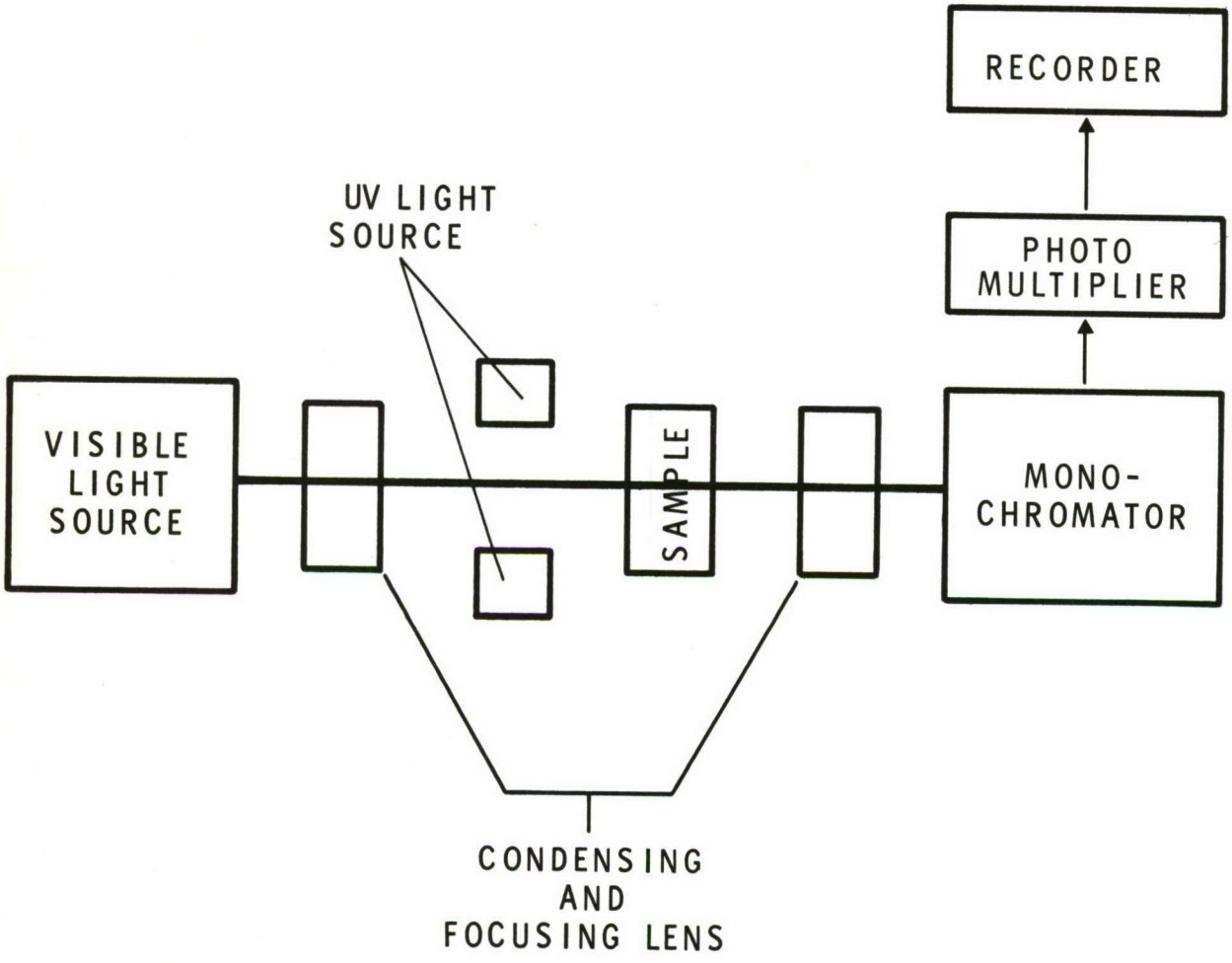


Figure 3. Flash Photolysis Unit

(U) A STUDY OF PANEL THICKNESS REQUIRED TO PREVENT FLUTTER IN ADVANCED PERFORMANCE VEHICLES

by

BOBBY L. LANGLEY, Capt., USAF

15 MAY 1964

Project Number 1370

Air Force Flight Dynamics Laboratory
Research and Technology Division
Air Force Systems Command
Wright-Patterson Air Force Base, Ohio 45433

ABSTRACT

The purpose of this study is to determine trends of panel thickness needed to prevent panel flutter for the Mach number ranges associated with several typical advanced vehicle trajectories and flight conditions. Also, the local flow conditions, such as local Mach number and local dynamic pressure, together with temperature, angle of attack and other parameters important to panel flutter are also presented and discussed.

The panel flutter parameters and trends are presented for several typical trajectories of advanced vehicles. A representative radiative corrugated stiffened panel was assumed to be located on the under side of the vehicles considered in this study. Panel flutter wind tunnel data available for the reference panel and extrapolations to higher Mach numbers are used together with the several advanced vehicles local flow conditions as determined by oblique shock theory to determine the trend of thickness required to prevent panel flutter.

Panels may experience local Mach numbers up to 24 for some configurations and flight conditions but, based on the results of this study, the most critical speed-range for panel flutter was generally in the transonic and low supersonic region of boosted flight. However, in some cases, the panel thickness to prevent flutter was nearly constant for free stream Mach number from 1.4 to 7.0.

LIST OF PRINCIPAL SYMBOLS

M_L - Local Mach Number

M - Free Stream Mach Number

P - Free Stream Pressure (psf)

P_L - Local Pressure (psf)

E - Modulus of Elasticity (psf)

t - Effective Thickness (in)

L - Length of Panel (in)

q_l - Local Dynamic Pressure (psf)

F - Flutter Parameter = $\left(\frac{E\beta}{q}\right)^{1/3}\frac{t}{L}$

$\beta = \sqrt{M^2-1}$

q_F - Flutter Dynamic Pressure (psf)

INTRODUCTION

Panel flutter has been of great concern since the development of aircraft with supersonic capabilities. This problem area will play a major role in the development program for future aerospace systems.

There has been considerable investigation of the flutter problems and panel flutter trends associated with a radiative corrugated stiffened panel (Fig. 1); therefore, this type panel was used as a basis for this study. This panel was assumed to be exposed to the flow and thermal environment for seven typical advanced vehicle flight paths and flow conditions. The thickness required to prevent flutter was calculated throughout the flight region.

Rene' 41 was the material utilized with the stiffness and load carrying properties of the reference panel were considered. However, Molybdenum alloy was used for vehicles that developed skin temperatures above about 1600° F. except for Vehicle #7. A Columbium alloy was used for Vehicle #7 since this material is under consideration due to characteristics of this vehicle and the associated trajectory.

Although a relatively simple type panel was used in this study, the corrugated type panel is the basis for advanced panels that are presently being considered for future vehicles. A corrugated panel with the addition of a heat shield seems to have qualities needed to meet the aerodynamic and structural requirements of tomorrow's high performance vehicles.

DISCUSSION OF PROCEDURES USED

The experimental flutter boundary curve for the reference panel used for this investigation was obtained from unpublished wind tunnel data for zero angle of attack for the Mach number range of 1.0 to 2.2 (Fig. 1). For Mach numbers greater than 2.2, the boundary was extrapolated by assuming that dynamic pressure continued to increase in a linear manner with increase in Mach number. The basis for this extrapolation is that according to theory, the dynamic pressure is proportional to $\sqrt{M^2-1}$ and for higher Mach numbers, $\sqrt{M^2-1}$ is approximately equal to M. Therefore, for higher Mach numbers, dynamic pressure is proportional to M. The flutter boundary above M = 2.2 (Fig. 1) is therefore a theoretical curve of dynamic pressure vs. M where dynamic pressure is proportional to Mach number and passes through the flutter point at M = 2.2.

While reliable experimental panel flutter trend information to higher Mach numbers is not available for this type of panel, some recent trend data (Ref. 1 & 2) for simple clamped panels to M = 6 indicates this linear extrapolation of the flutter boundary with Mach number is reasonable. Data from Ref. 1 and 2 was reviewed and a definite linear trend was present as predicted by theory for the Mach number range above 2. This linear trend was present even though the slopes of the q vs. M curves for the experimental panels differed from the slope of the reference panel curve. These different slopes were associated with panels of various panel thickness, material and type construction.

The effective thickness to prevent flutter for this study was determined from the previously developed flutter boundary (Fig. 1) and also appropriate relationships based on the panel flutter parameter:

$$F = \left(\frac{E\sigma}{q}\right)^{1/3} \frac{t}{L}$$

First, the parameter

$$F_1 = \left(\frac{E_1 \sigma_1}{q_1}\right)^{1/3} \frac{t_1}{L_1} \tag{1}$$

pertains to the advanced vehicle panel local flight conditions based on the trajectories and local Mach number, local dynamic pressure, and modulus of elasticity for the panel temperatures. Second, the parameter,

$$F = \left(\frac{E_2 \sigma_2}{q_2}\right)^{1/3} \frac{t}{L} \tag{2}$$

pertains to the panel flutter boundary conditions for the reference panel as presented (Fig. 1) and based on wind tunnel flutter data with analytical extrapolations to higher Mach numbers. The flutter boundary for the panel with the advanced vehicle trajectory is defined when the vehicle panel local conditions equal the panel flutter boundary conditions or $F_1 = F_2$. Therefore,

$$\left(\frac{E_1 \sigma_1}{q_1}\right)^{1/3} \frac{t_1}{L_1} = \left(\frac{E_2 \sigma_2}{q_2}\right)^{1/3} \frac{t_2}{L_2}$$

If the vehicle panel local condition parameter, F_1, was less than the panel flutter boundary condition, F_2, flutter would occur. If F_2 was greater than F_1, flutter would not occur. The nondimensional thickness ratio required to flutter on the representative panel for the advanced vehicle trajectory is

$$\frac{t_1/L_1}{t_2/L_2} = \left(\frac{\sigma_2 E_2 q_1}{\sigma_1 E_1 q_2}\right)^{1/3} \tag{3}$$

or for $L_1 = L_2$ (same length of panel for this study as used in the wind tunnel experiments) and with a given effective thickness, t_2, for the panel used in the experimental flutter boundary, the thickness to prevent flutter for the advanced vehicle trajectory can be determined as:

$$t_1 = t_2\left(\frac{\beta_2 E_2 q_1}{\beta_1 E_1 q_2}\right)^{1/3} \tag{4}$$

Further, since the required thickness was determined for discrete local Mach numbers based on the wind tunnel flutter data, for the same Mach number, $\beta_1 = \beta_2$ and

$$t_1 = t_2\left(\frac{E_2 q_1}{E_1 q_2}\right)^{1/3} \tag{5}$$

Equation (5) together with the wind tunnel data in Figure 1 and the flight trajectory data for the advanced vehicles are the basis for this study to determine the thickness required to prevent flutter versus free stream Mach number for the several advanced vehicles.

Data was computed for all flight conditions of varying angles of attack, Mach number, altitude and skin temperature for the seven trajectories considered. The temperatures used were those calculated for an uninsulated structure at any given time during the trajectory. All the above variables were determined from predicted performance data. With the given parameters, oblique shock theory was used to determine values for local conditions behind the shock. These values were used in equation 5 to determine the thickness required.

The reference panel is of corrugated structure and has been calculated to have an effective thickness of 0.01235" (Ref. 3). The reference panel utilized corrugations normal to the stream direction,

thus compensating for high temperature gradients. By using the corrugation in the streamwise direction, the effective thickness of the panel would be increased 15 times (Ref. 4) and be more satisfactory from a flutter prevention standpoint. Any mention of thickness during the remainder of this report will imply effective thickness of an isotropic panel with the same effective stiffness as the reference built-up panel.

RESULTS

VEHICLE #1

Fig. 2 shows the local flow conditions and Fig. 3 shows the thickness required to prevent flutter for vehicle #1. During exit, the local Mach number, M_L, varied from subsonic to 7.2 (M_∞ = 24.2) near the end of the exit phase (Fig. 2). The angle of attack varied from 1 to 14°. The local dynamic pressure, q_L, for the exit phase was greater than any other trajectory reviewed. The maximum q_L was 4250 psf and occurred at M_∞ = 7 (M_L = 5.3). Fig. 3 shows the maximum effective thickness required to prevent flutter occurring at M_∞ = 4.2 (M_L = 3.8) where a near maximum in q_L occurred. This thickness requirement (Fig. 3) was nearly constant from transonic velocities to M_∞ = 7. The range of M_∞ above 12 was associated with very high altitudes causing q_L and q_∞ to be very low and thus the thickness requirement to be small.

The re-entry trajectory for this vehicle is the same as for the Vehicle #4 trajectory which is presented in Fig. 11 and discussed in a later section. The maximum thickness required during re-entry was about 52% of the maximum required during exit.

VEHICLE #2

Fig. 4 shows the local flow conditions and Fig. 5 shows the thickness required to prevent flutter for Vehicle #2. The M_L varied from subsonic to 24 (M_∞ = 24) during the exit phase, (Fig. 4). The extensive range of M_L was due to small angles of attack (0 to 5°). The maximum q_L (Fig. 4) was 2700 psf and occurred at M_∞ = 9 (M_L = 8). The maximum

thickness (Fig. 5) required to prevent flutter was at $M_\infty = 1.6$ ($M_L = 1.4$). At $M_\infty = 1.6$ the thickness required reduced rapidly with increasing Mach number (thickness at $M_\infty = 6$ is about 30% of that at $M_\infty = 1.6$); then the thickness continued to decrease slowly with increasing Mach number to about $M_\infty = 22$ where the thickness was about 50% of the maximum at transonic speeds. The re-entry trajectory is not available for this system.

VEHICLE #3

Fig. 6 shows the local flow conditions for exit and Fig. 7 shows the thickness required to prevent flutter for Vehicle #3. The M_L varied from subsonic to 22.6 during the exit phase (Fig. 6). This large range of local Mach numbers was caused by relatively low angles of attack (0 to 5°) at high free stream Mach number. The local dynamic pressure, q_L, is nearly constant from $M_\infty = 3$ to 10 ($M_L = 3$ to 9) with the maximum of 2600 psf at $M_\infty = 5.15$ ($M_L = 4.4$). At M_∞ greater than 9, q_L decreases rapidly. The maximum thickness to prevent flutter (Fig. 7) occurred at about $M_\infty = 2$ ($M_L = 1.9$). For lower or higher M_∞, the thickness required to prevent flutter was rapidly reduced.

The maximum thickness required during re-entry occurred at $M_\infty = 14$ ($M_L = 4.6$). Fig. 8 shows the local flow conditions for Vehicle #3 during re-entry. During re-entry the q_L peaked at 740 psf at $M_\infty = 9$ ($M_L = 4.8$). The local Mach number stayed below 5.0 during re-entry because of higher angles of attack (12-24°) than during exit. Fig. 7 indicates that the re-entry condition should be much less severe than the exit phase from the panel flutter standpoint.

VEHICLE #4

Fig. 9 shows the thickness required to prevent flutter and Fig. 10 and Fig. 11 show the local flow conditions for Vehicle #4. In the exit phase, the maximum thickness required to prevent flutter (Fig. 9) occurred at $M_\infty = 2$ ($M_L = 1.9$) and declined only slightly to $M_\infty = 4$. At M_∞ above 4, the thickness to prevent flutter decreased rapidly with increasing Mach number. At $M_\infty = 2$, where thickness required is a maximum with the vehicle at 5° angle of attack, the local dynamic pressure and free stream dynamic pressure were approximately equal with values of 1400 psf. At angles of attack of 5°, the decrease in M_L and increase in P_L were of the right magnitude to equal the combination of M_∞ and P_∞ when dynamic pressures were computed. The local dynamic pressure peaked at $M_\infty = 4.15$ and a value of 3000 psf. As the altitude increased after this peak, the pressure decreased, causing a rapid decrease in dynamic pressure. M_L and M_∞ (Fig. 10) remained approximately equal throughout the trajectory due to low angles of attack (1° to 5°) and had a range of subsonic to 23.

The re-entry phase (Fig. 11) covered a M_L range of 5.15-12.0 ($M_\infty = 7$-27). At an altitude of 250,000 feet, the constant angle of attack of 9° produced a $M_L = 12$ for $M_\infty = 27$. At lower M_∞, this difference was not so prominent. For example, at an angle of attack of 9° and $M_\infty = 3$, the resulting M_L is 2.7. The maximum q_L during re-entry was 865 psf and occurred at $M_\infty = 11$ ($M_L = 7.3$). The maximum thickness to prevent flutter during re-entry (Fig. 9) occurred at $M_\infty = 10$ ($M_L = 7$). This

thickness requirement was approximately constant from M_∞ = 10-16. The maximum thickness required during re-entry was approximately one-half that required during the initial low supersonic portion of exit. However, at free stream Mach number greater than about 9, the re-entry flight required much more thickness to prevent flutter than the exit portion corresponding to the same free stream Mach numbers.

VEHICLE #5

Fig. 12, Fig. 13, and Fig. 14 show the local flow conditions and Fig. 15 shows the thickness required to prevent flutter for Vehicle #5. The boost trajectory used was for a constant angle of attack of 5°. The range of M_L (Fig. 12) for the boost phase was from subsonic to 2.7 (M_∞ = 3). Data was not available above M_∞ = 3. The maximum q_L for boost was 110 psf and occurred at M_∞ = 2 (M_L = 1.8). The maximum thickness to prevent flutter (Fig. 15) occurred at M_∞ = 1.5 (M_L = 1.32). As can be seen in Fig. 15, the maximum thickness required to prevent flutter is well below the thickness of the reference panel and if the reference panel was used, flutter would not occur. If a panel with an outer skin of .007" were used instead of the .008" thickness of the reference panel, flutter would occur. From equation (5) it can be seen that the flutter dynamic pressure varies as the cube of the effective thicknesses thus $q \propto t^3$. From the above relationship, it can be seen that a 33.5% decrease in flutter dynamic pressure is brought about by only 12.7% decrease in effective thickness. The minimum Rene' 41 outer skin thickness to prevent flutter is .00708". This tends to point out the extreme sensitivity of the dynamic pressure margin of safety to very small

changes in skin thickness.

Of the three re-entry trajectories for this vehicle, the maximum (Fig. 13) was 662 psf at M_∞ = 8.1 (M_L = 4.5) for the maximum lateral range. The most critical panel flutter conditions for re-entry was at q_L of 473 psf at M_∞ = 1.65 (M_L = 1.47) for the Maximum Longitudinal Range (Fig. 14). The maximum thickness required is .0086" at M_∞ = 1.65 (M_L = 1.47). The boost phase was the most critical phase and the maximum thickness to prevent flutter was 22% greater than during the re-entry phase.

VEHICLE #6

Fig. 16 and Fig. 17 show the local flow conditions and Fig. 18 shows the thickness required to prevent flutter for Vehicle #6. The boost phase covers a range of M_L from subsonic to 8.0 (M_∞ = 9.2) as shown in Fig. 16. A maximum q_L of 1000 psf occurred at M_L = 1.50 (M_∞ = 2). The peak in q_L was followed by lower values at higher Mach numbers. The maximum thickness required (Fig. 18) occurred at M_∞ = 2.0 (M_L = 1.55). M_∞ below 1.6 was not computed due to the shock theory (Ref. 5) being invalid at the low Mach numbers and high angles of attack for that condition.

During the re-entry, the M_L decreased from 3.8 (M_∞ = 11.8) to 1.43 (M_∞ = 4.5). Data is not available for lower Mach number. This large difference between M_∞ and M_L was brought about by the high angle of attack. The angle of attack varied from 31° to 37.5°. M_L stayed low and at one point (M_∞ = 9.5) was only 22% of M_∞. The thickness to prevent flutter remained approximately constant during the glide (Fig. 18). This was about 30% lower than the maximum thickness required to prevent panel flutter during boost.

VEHICLE #7

Fig. 19 shows the thickness required to prevent flutter for both exit and re-entry. Fig. 20 and 21 show the local flow conditions for both exit and re-entry. This trajectory was the only one of the seven reviewed that produced panel flutter conditions (the thickness required to prevent flutter greater than the thickness of the reference panel). This was brought about largely by the low (16.2×10^6 psi as compared to 31.5×10^6 psi for the Rene' 41 reference panel) modulus of elasticity of the Columbium alloy. The maximum thickness required (Fig. 19) during exit occurred at $M_\infty = 2.4$ ($M_L = 2.15$). Maximum q_L was 3250 psf and occurred at $M_\infty = 7.7$ ($M_L = 6.2$).

The maximum thickness required to prevent flutter for this trajectory occurred during the re-entry phase. This maximum occurred at $M_\infty = 1.5$ and $q_L = 1190$ psf. The maximum q_L (Fig. 21) was 2560 psf at $M_\infty = 4.6$. This was the only trajectory for which the maximum thickness required to prevent flutter occurred during the re-entry phase and also the only trajectory that had thickness requirements that exceeded the effective thickness of the reference panel. If the reference panel were used, flutter would occur.

The flutter condition associated with Vehicle #7 was brought about mainly by the low value of the modulus of elasticity of the Columbium alloy. The dynamic pressures associated with this trajectory did not differ greatly from those of other vehicles considered. It can be seen from Table 1 that the ratio of the reference panel modulus of elasticity to the vehicle panel modulus of elasticity $\frac{E_2}{E_1}$ is very high compared to

the corresponding values associated with the other trajectories. This indicates that the Columbium alloy has lower stiffness properties than the reference panel and thus increases the panel thicknesses needed to prevent flutter.

Table 1 is a summary of the parameters and ratios that affect the panel thickness to prevent flutter. 'Although re-entry data is limited in most cases to altitudes above 60,000 ft., it can be seen that the maximum thickness required during re-entry does not always occur at the minimum M_∞ computed for the trajectory. In general, M_∞ decreases with a decrease in altitude during the final stages of re-entry. This indicates that thickness requirements do not always increase with a decrease in altitude as in the case of Vehicle #7. Although, the very low altitudes are the most critical for Vehicle #7, this trend is not present for all vehicles.

Table 2 is a summary of the maximum temperatures for both exit and re-entry for the seven vehicles. In all cases, the re-entry maximum temperature is greater than the exit.

In addition, a superorbital trajectory was considered and Newtonian theory (Ref. 6) was used to compute the pressure coefficients for the vehicle. Local dynamic pressure was found to be extremely low (190 psf) even at $M_\infty = 26.7$ where the vehicle dips into the atmosphere at 195,000 ft. The bluntness of the body and the high re-entry angle of attack made the oblique shock theory invalid. The extremely high

temperatures associated with this trajectory make the use of ablative materials necessary and thus the trajectory was considerably different than the other seven presented. The use of ablative materials would tend to increase the stiffness properties of the panels. The combination of low local dynamic pressure and increased stiffness would point to a minimum of panel flutter difficulties in this area of high altitude and velocities.

CONCLUSIONS

The following conclusions are based on the configurations, flow conditions and trajectories presented herein.

1. The maximum q_L was always greater in the boost (exit) phase than for re-entry and in all cases, except in Mach number range of M_∞ = 1-1.5 of the trajectory for Vehicle #5, q_L was greater than q_∞.

2. Some of the trajectories studied for advanced aerospace designs involved small angles of attack (under 14°) throughout the flight. These small angles of attack gave local Mach numbers nearly equal to free stream Mach numbers and up to 24 on the lower surface of the vehicle.

3. Maximum panel temperature was always greater for the re-entry phase than for the boost phase (Table 2).

4. In all cases considered, except Vehicle #7, the maximum thickness required to prevent flutter occurred during the boost (exit) phase rather than for the re-entry.

5. The maximum thickness required to prevent flutter during exit was found in the M_∞ range of 1.5-2.5 for all cases considered except for the Vehicle #1 trajectory. In this case, the maximum thickness required remained high over a range from about 1.4-7.0.

6. The condition corresponding to maximum q_L was not always the point of maximum thickness required to prevent flutter. This occurs because high q_F are associated with high Mach numbers as shown in Fig. 1. The thickness required to prevent flutter is proportional to $\left(\frac{q_L}{q_F}\right)$. Therefore, if the maximum q_L occurs at high M_L, the corresponding high

q_F tends to minimize the effect of the maximum q_L when computing the thickness to prevent flutter.

7. It appears that for panels of corrugated construction as considered herein, small skin thickness changes (.001") make significant alterations in panel flutter characteristics. A decrease in skin thickness from .008" to .007", reduced the effective thickness by 12.7% and reduced the flutter dynamic pressure by 33.5%.

8. The flutter condition associated with Vehicle #7, which was the only vehicle to use Columbium alloy, was brought about by the low value of the modulus of elasticity of the Columbium alloy.

RECOMMENDATIONS

1. More complete re-entry data is needed, particularly for altitudes below 60,000 feet. This could be a problem area, but limited data prevented a complete study at this time and should be considered further when such data become available.

2. A more comprehensive study including panel configurations representative of those for the actual vehicle under consideration and for various specific locations on vehicles should be conducted when more data are available on the advanced vehicles.

3. Another area of interest is the weight efficiency of the advanced vehicle panel construction. The feasibility of constructing a panel to withstand static and dynamic loads and still be sufficiently lightweight, falls in this area. The lack of definitive information on materials and insulation methods to be used made this study very difficult at this time.

4. The results for both re-entry and exit for high Mach number conditions were highly dependent on the accuracy of the theoretical extrapolation of the experimental flutter curve above M = 2.2. The confirmation of this curve with further hypersonic experimental data would be worthwhile and necessary to form positive conclusions. Further hypersonic panel flutter data should be obtained by wind tunnel or flight test to establish and verify hypersonic prediction methods. Some flight tests are presently being planned as part of the ASSET re-entry program.

5. Thermal effects on panel flutter during re-entry were treated only as a reduction in modulus of elasticity of the material and thus a reduction in the effective panel stiffness. Since in-plane stresses induced by increased temperature tend to decrease panel vibration frequencies and lower the flutter dynamic pressure, further work to better identify these effects would be worthwhile.

REFERENCES

1. E. D. Dowell and H. M. Voss, "Experimental and Theoretical Panel Flutter Studies in the Mach Number Range of 1.0 to 5.0," ASD-TDR-63-449, Aeronautical System Division, Wright-Patterson AFB, Ohio, Dec 1963 (Confidential Report)

2. D. J. Ketter and H. M. Voss, "Panel Flutter and Experiments in the Mach Number Range of 5.0 to 10.0," FDL-TDR-64-6, Flight Dynamics Laboratory, Wright-Patterson AFB, Ohio, Jan 1964 (Confidential Report)

3. William J. Bozich, Capt. USAF, "The Effective Stiffness and Thickness of Corrugation-Stiffened Panels," ASRMD-TM-63-5, Aeronautical System Division, Wright-Patterson AFB, Ohio, May 1963

4. W. Jefferson Stroud, "Elastic Constants for Bending and Twisting of Corrugation-Stiffened Panels," NASA TR R-166, Langley Research Center, Langley Station, Hampton, Va.

5. National Advisory Committee for Aeronautics, Report 1135, Equations, Tables and Chart for Compressible Flow, Ames Research Staff, 1953

6. Valentine Dahlam III, "Hypersonic Aerodynamic Characteristics of Arbitrary Bodies," ASRMDF-TM-62-61, Sept 1962

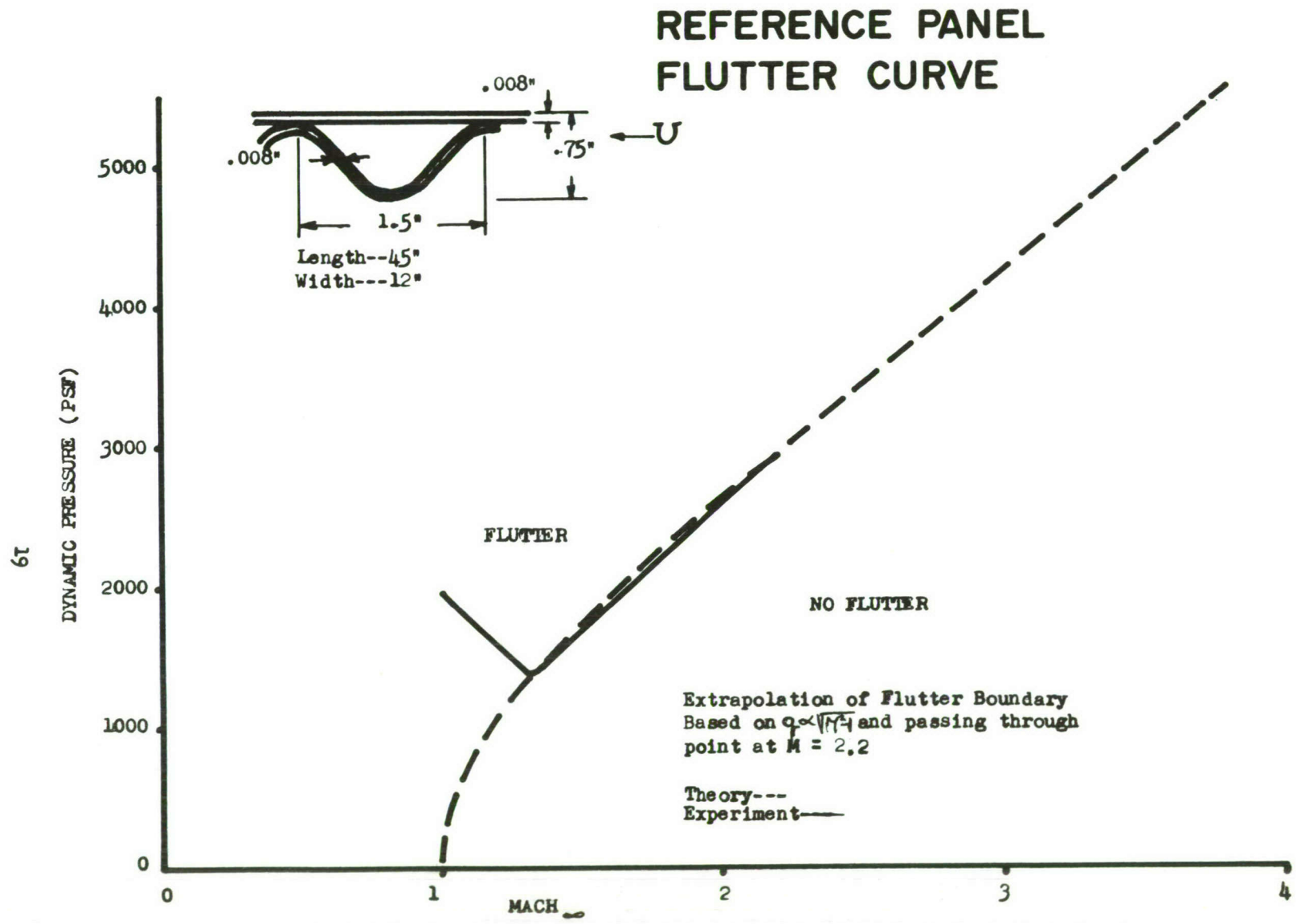


Fig. 1 Mach No. vs Dynamic Pressure for Reference Panel

VEHICLE NO. 1
MOLYBDENUM PANEL
EXIT

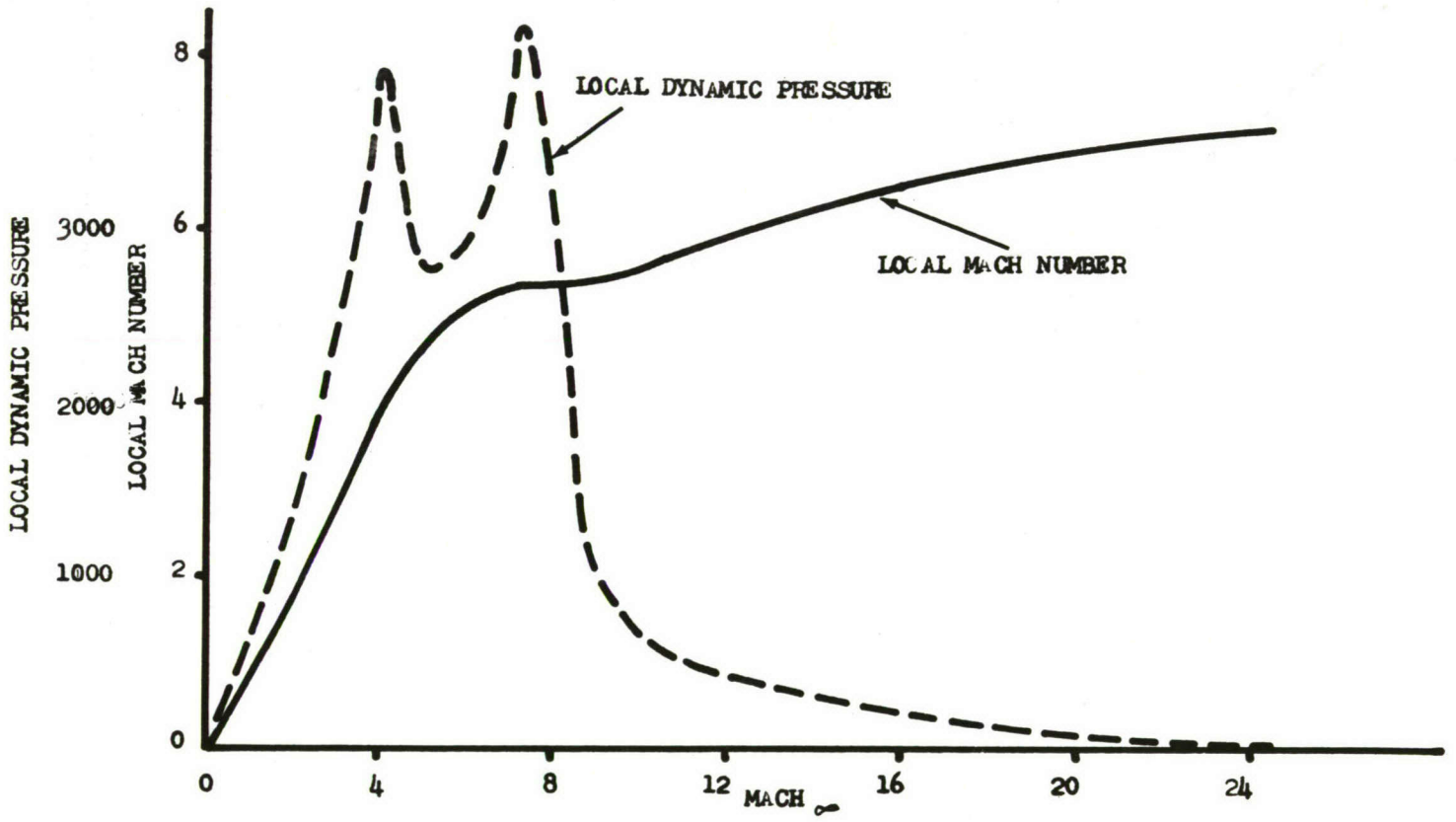


Fig. 2 Mach No. vs Local Flow Conditions (Veh. #1)

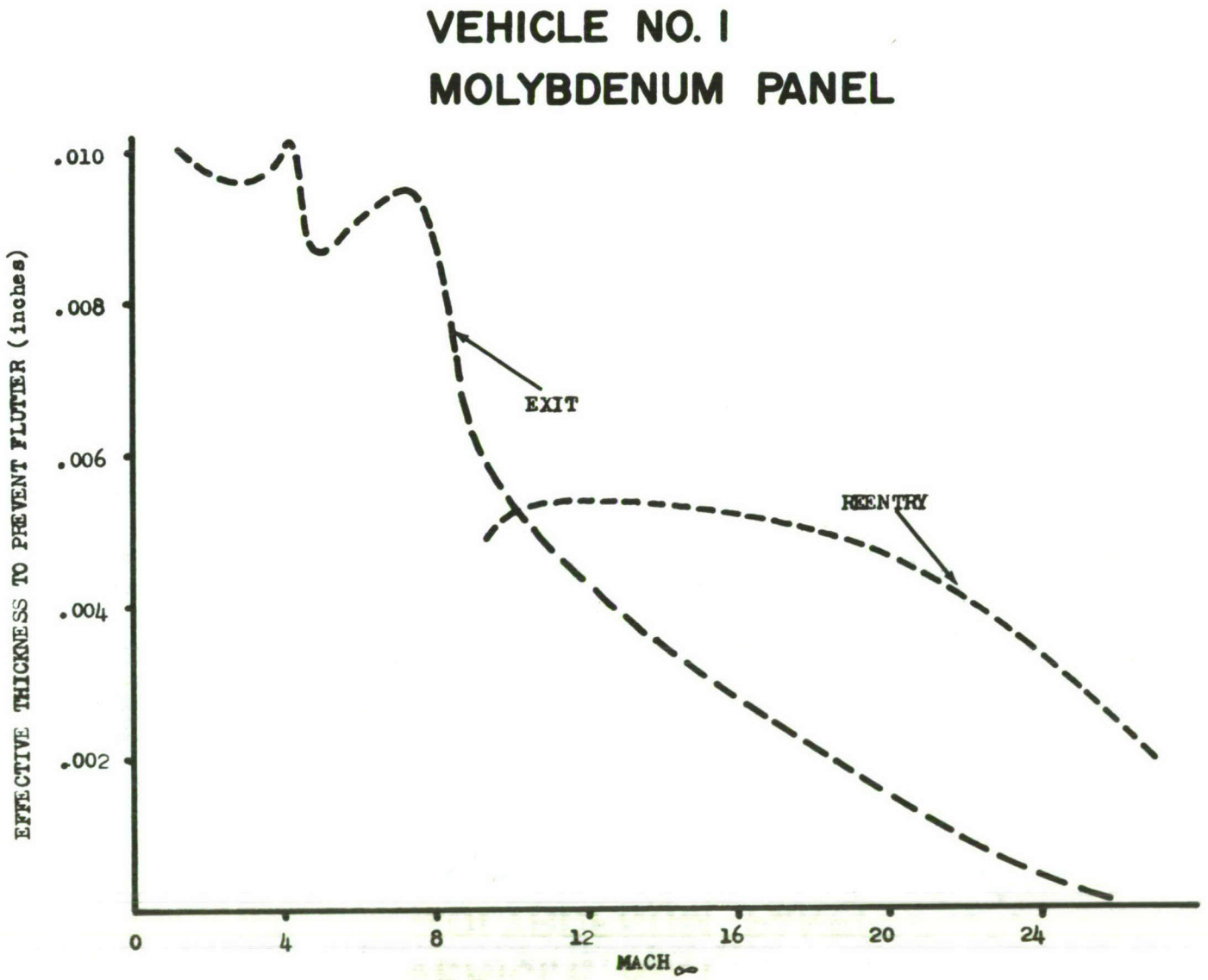


Fig. 3 Mach No. vs Thickness Required to Prevent Flutter (Veh. #1)

VEHICLE NO. 2
MOLYBDENUM PANEL

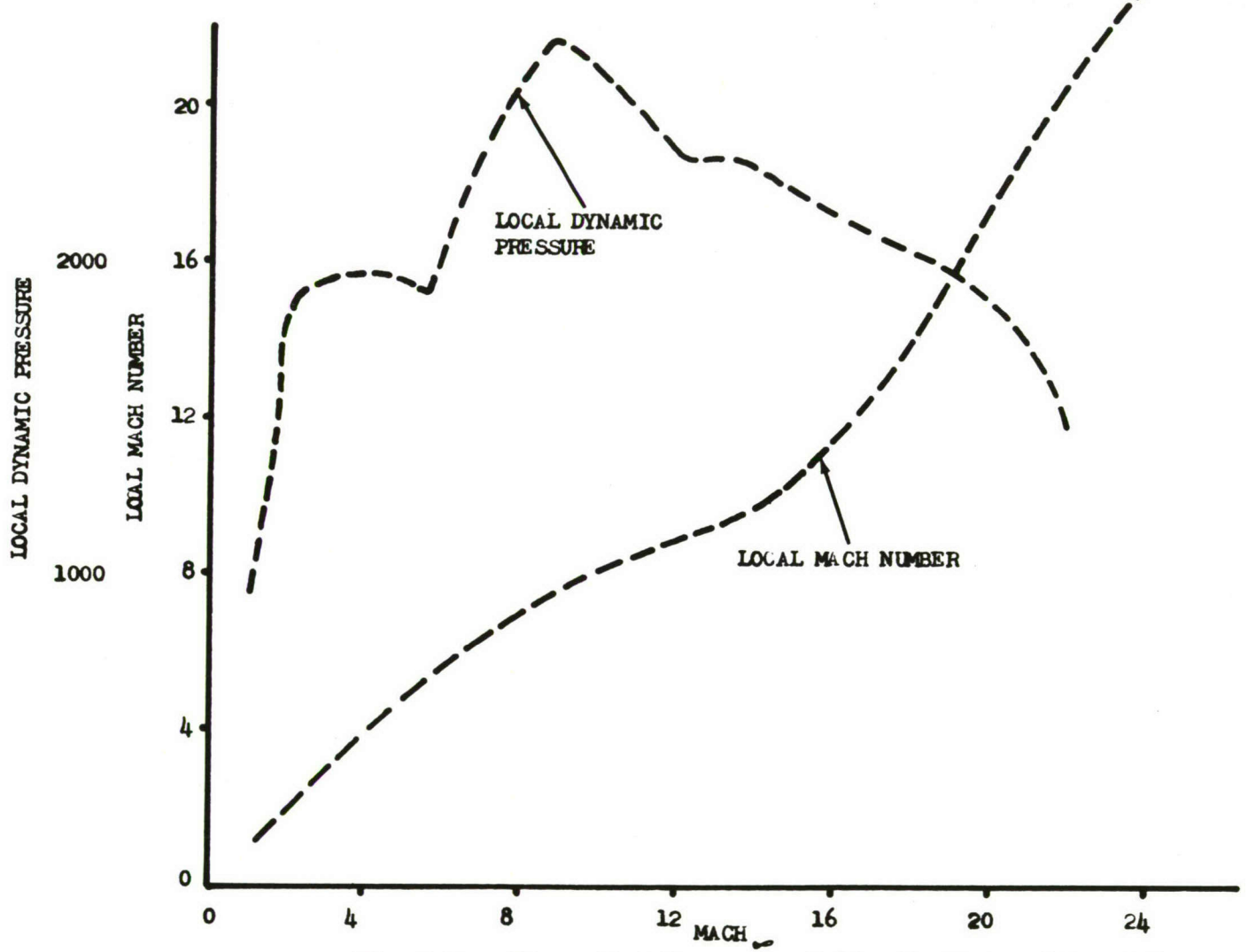


Fig. 4 Free Stream Mach No. vs Local Flow Conditions (Veh. #2)

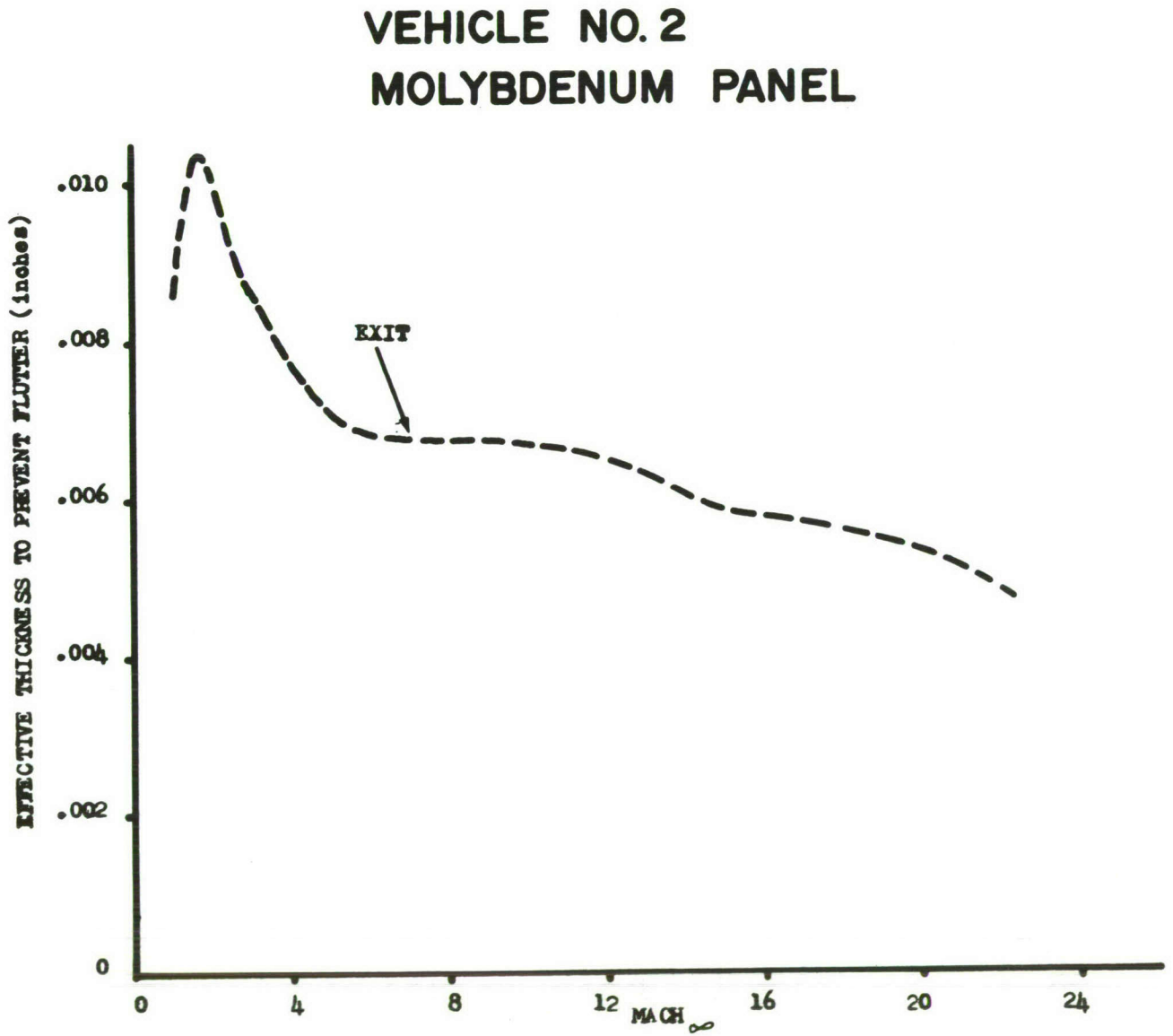


Fig. 5 Mach No. vs Thickness Required to Prevent Flutter (Veh. #2)

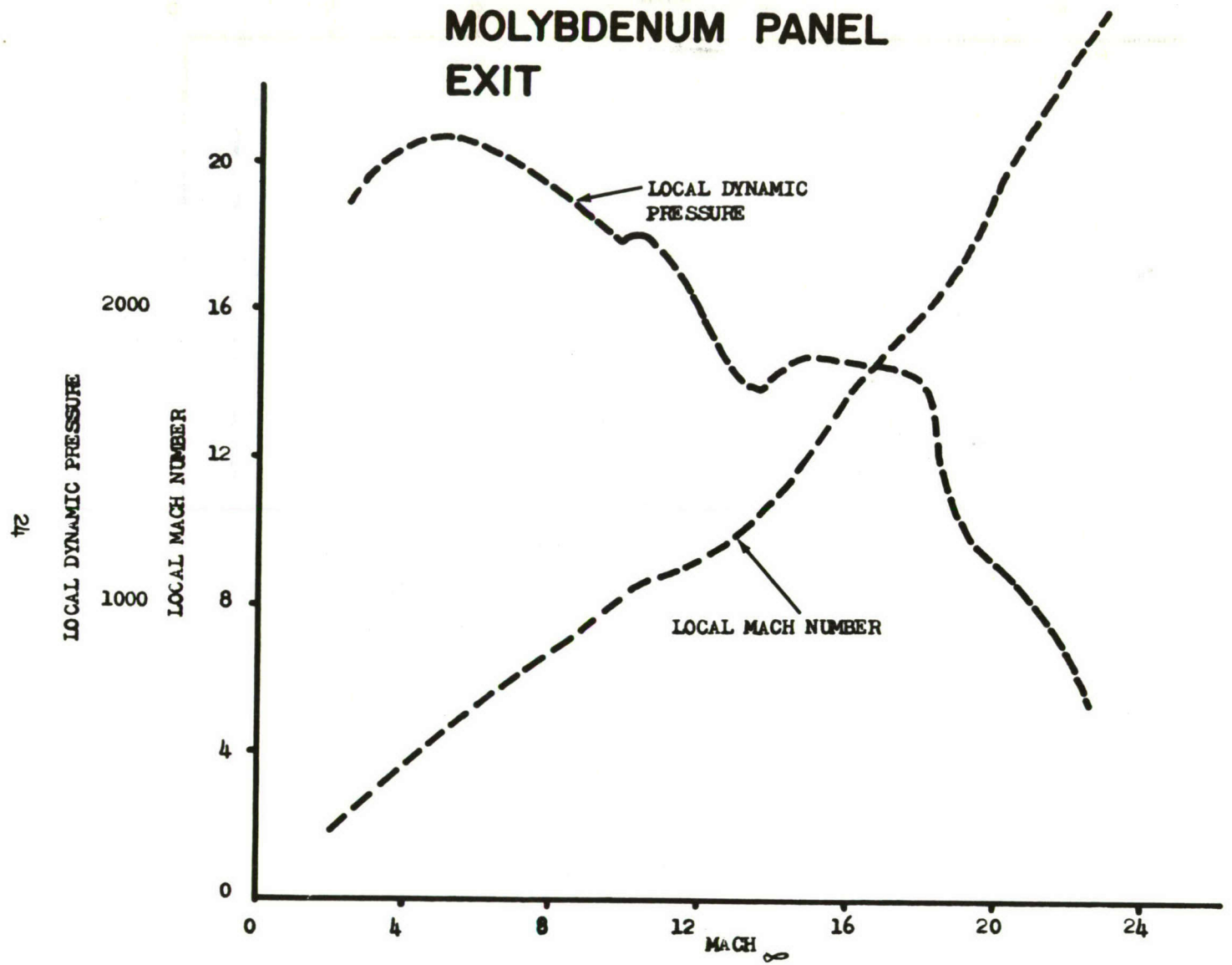


Fig. 6 Free Stream Mach No. vs Local Flow Conditions (Veh. #3)

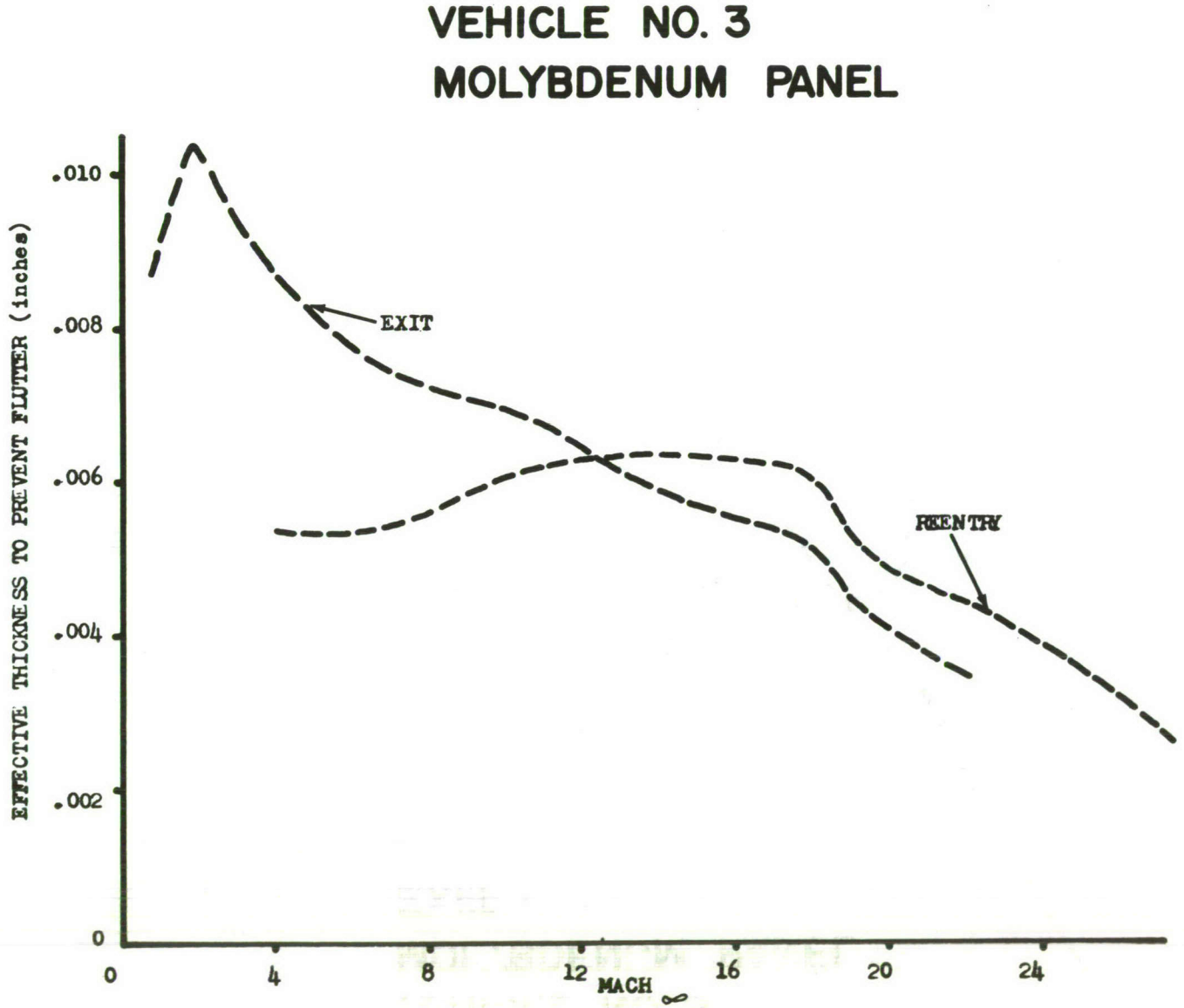


Fig. 7 Mach No. vs Thickness Required to Prevent Flutter (Veh. #3)

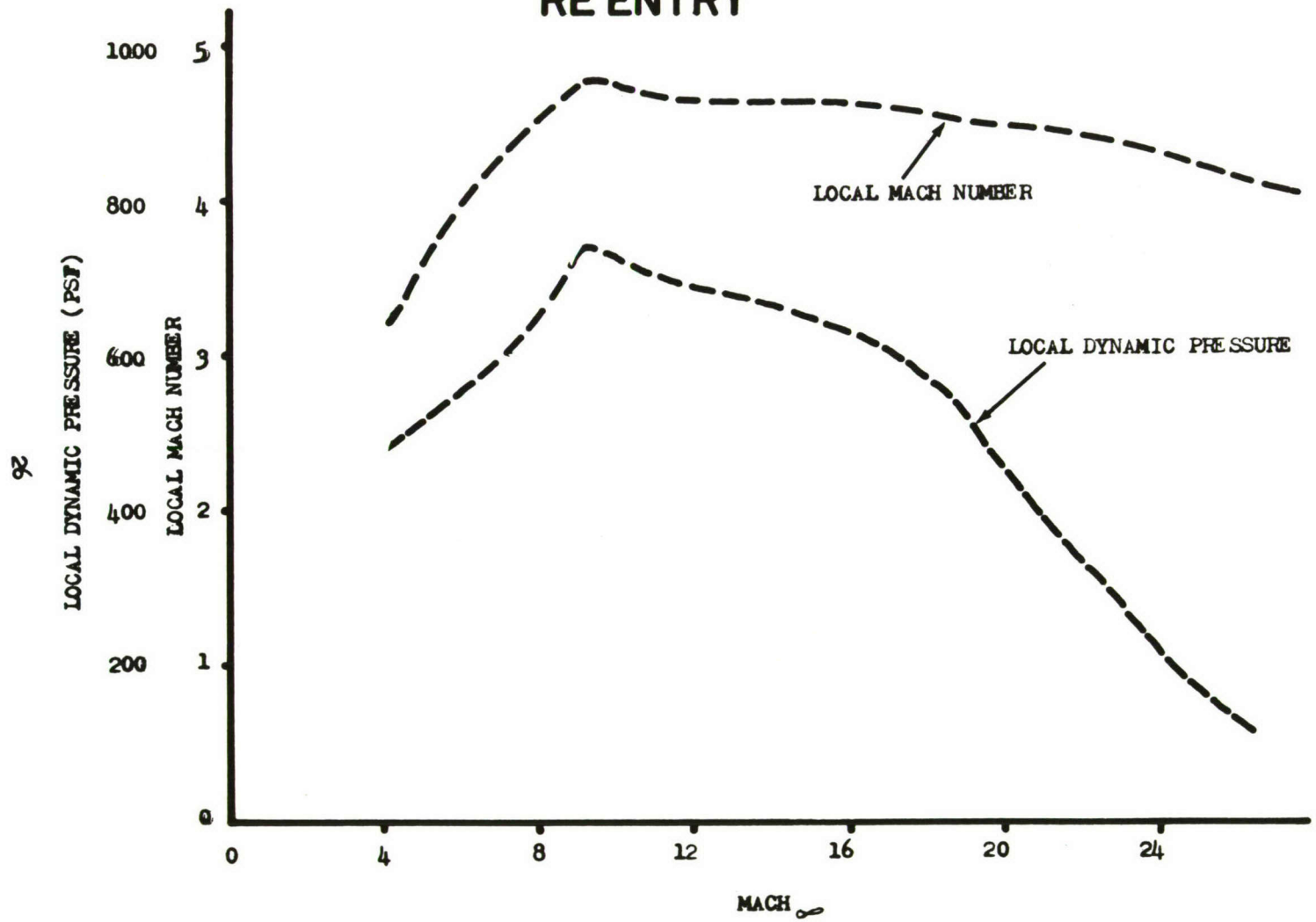


Fig. 8 Free Stream Mach No. vs Local Flow Conditions (Veh. #3)

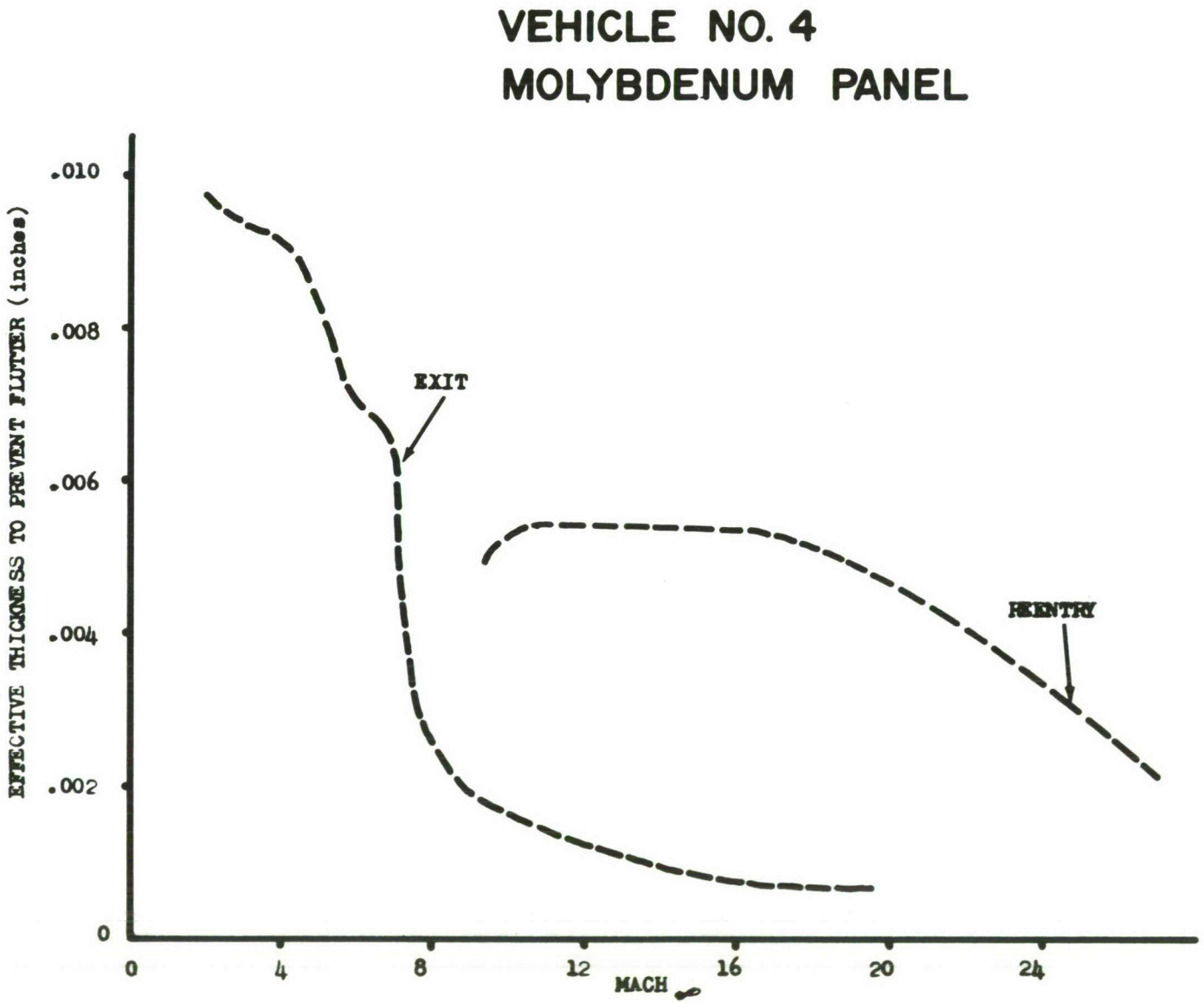


Fig. 9 Mach No. vs Thickness Required to Prevent Flutter (Veh. #4)

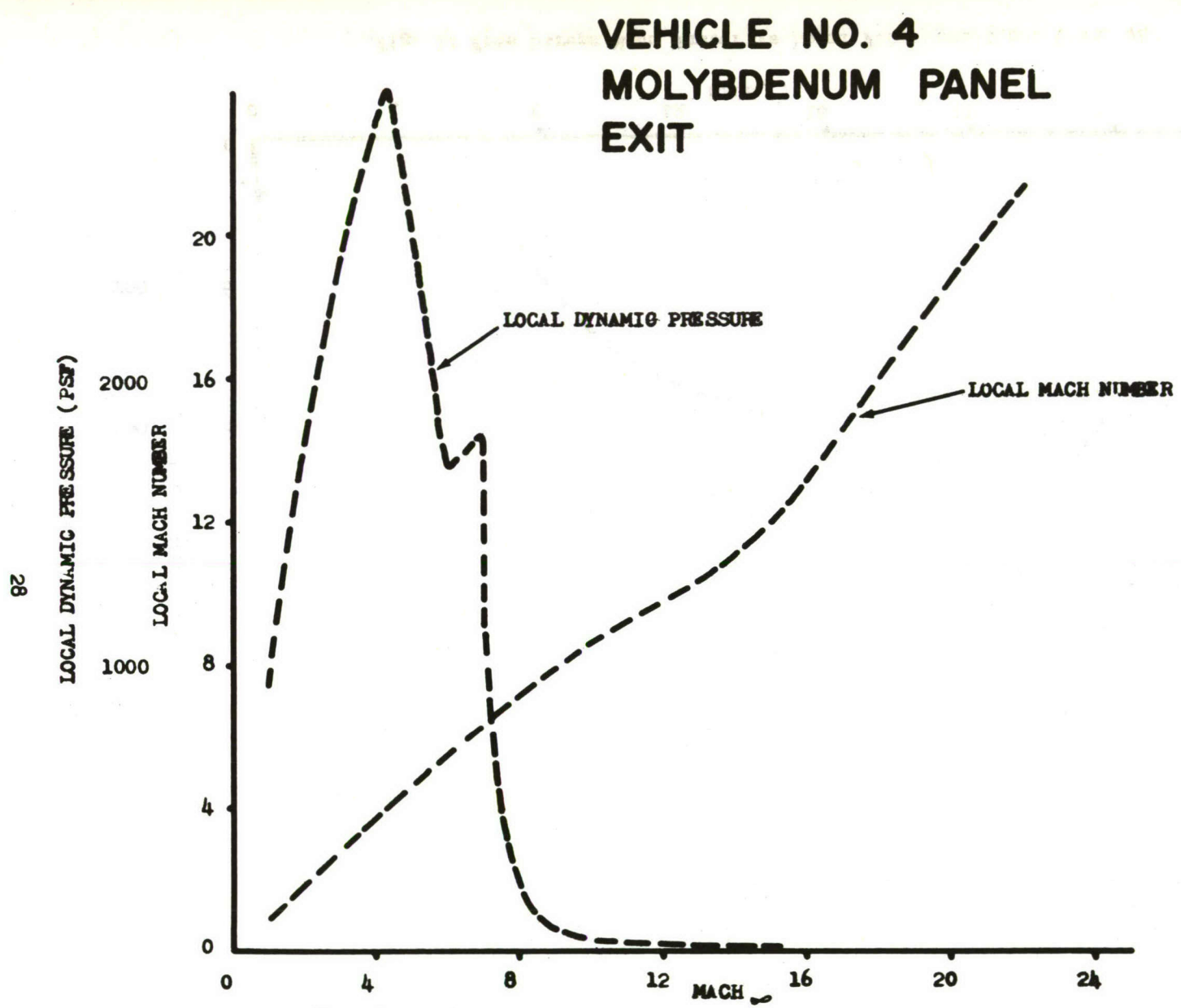


Fig. 10 Free Stream Mach No. vs Local Flow Conditions (Veh. #4)

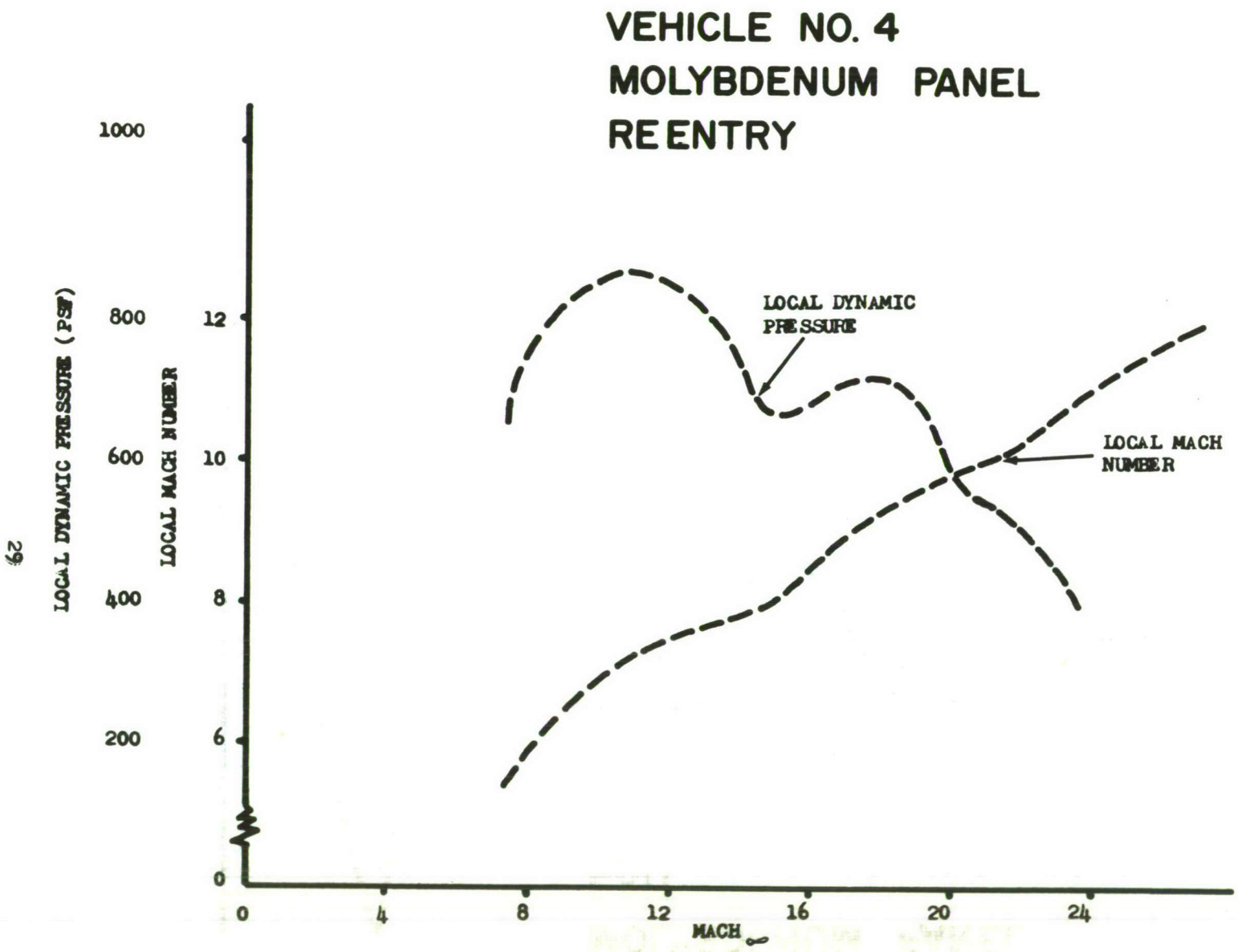


Fig. 11 Free Stream Mach Number vs Local Flow Conditions (Veh. #4)

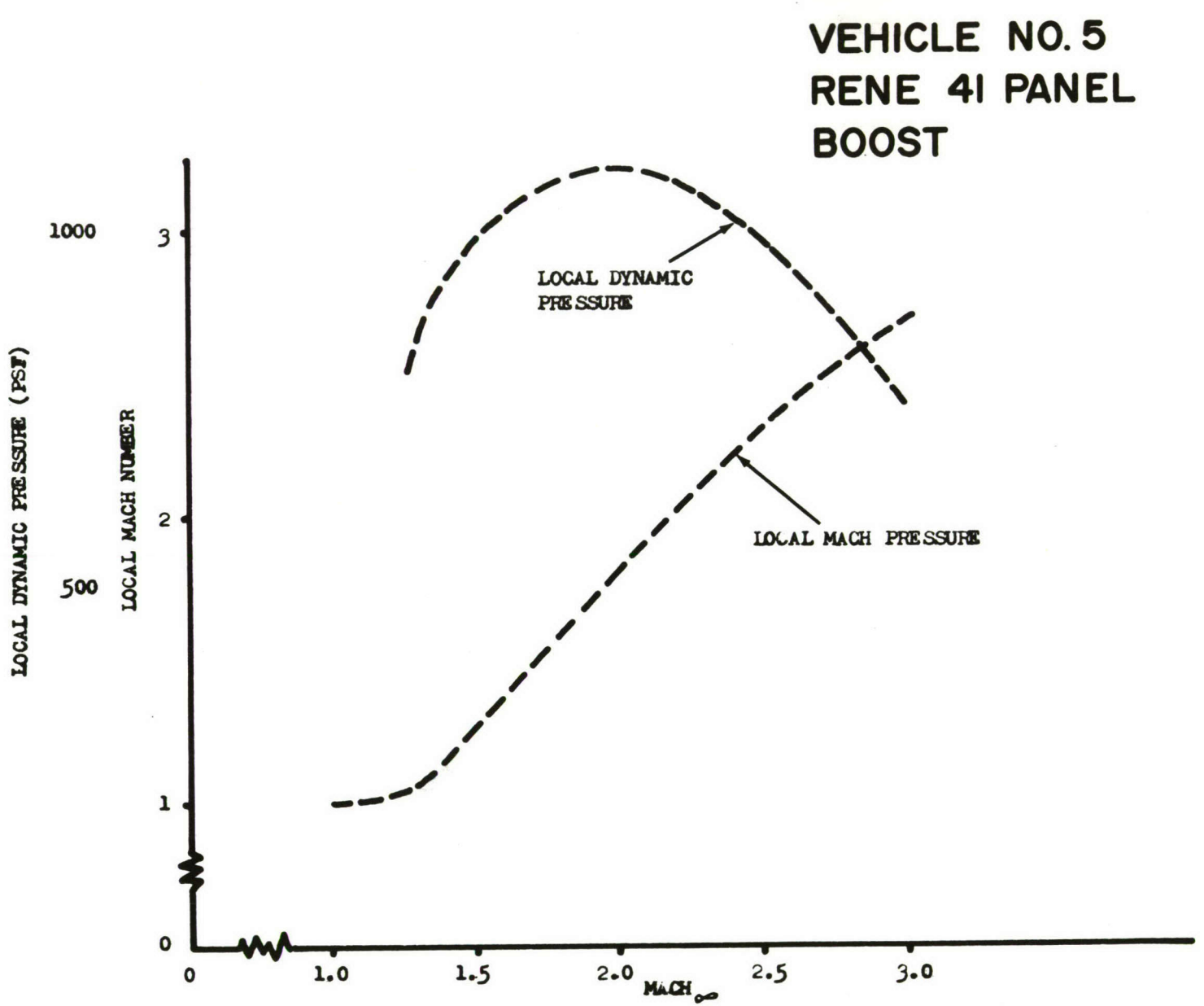


Fig. 12 Free Stream Mach No. vs Local Flow Conditions (Veh. #5)

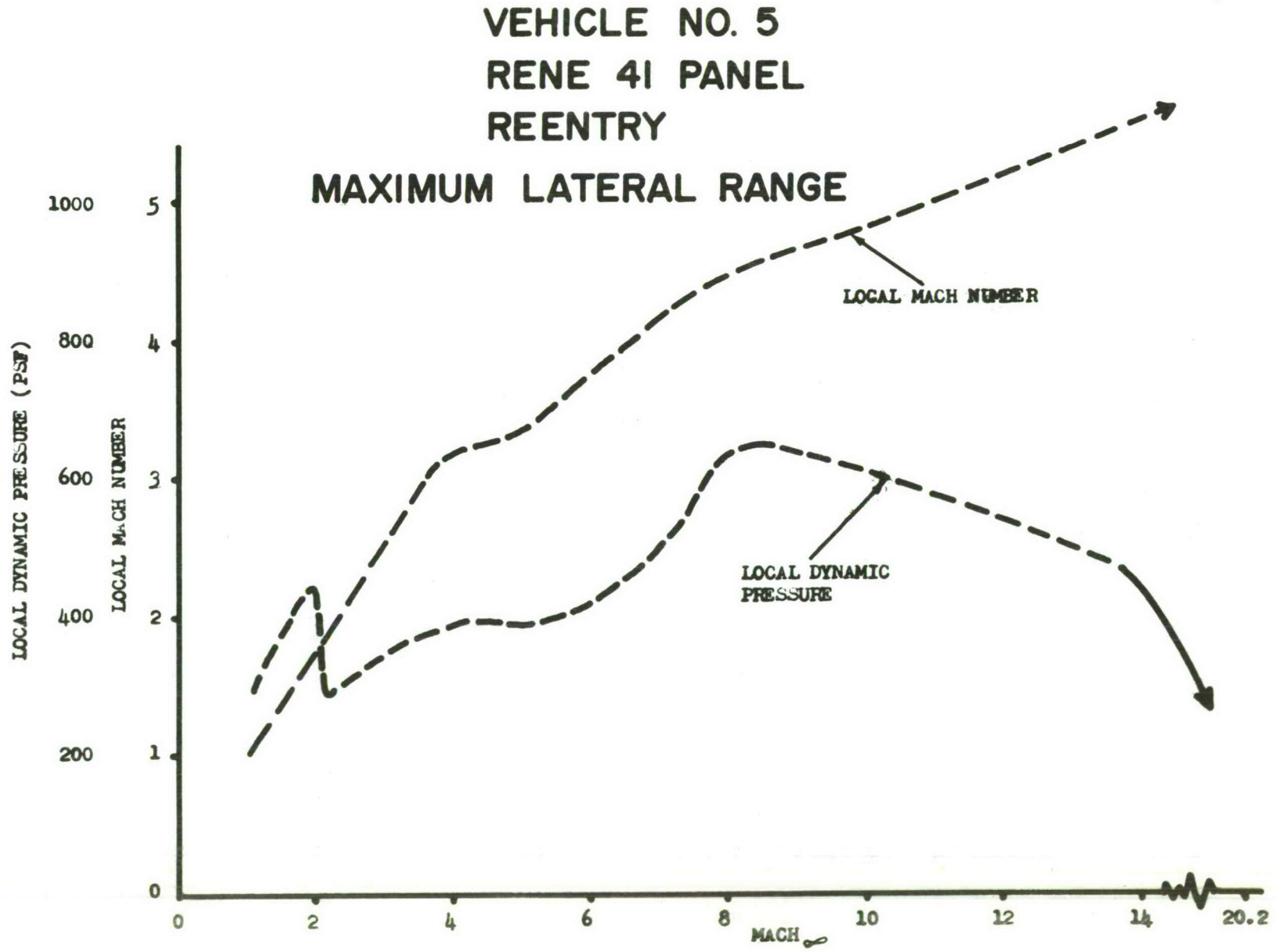


Fig. 13 Free Stream Mach No. vs Local Flow Conditions (Veh. #5)

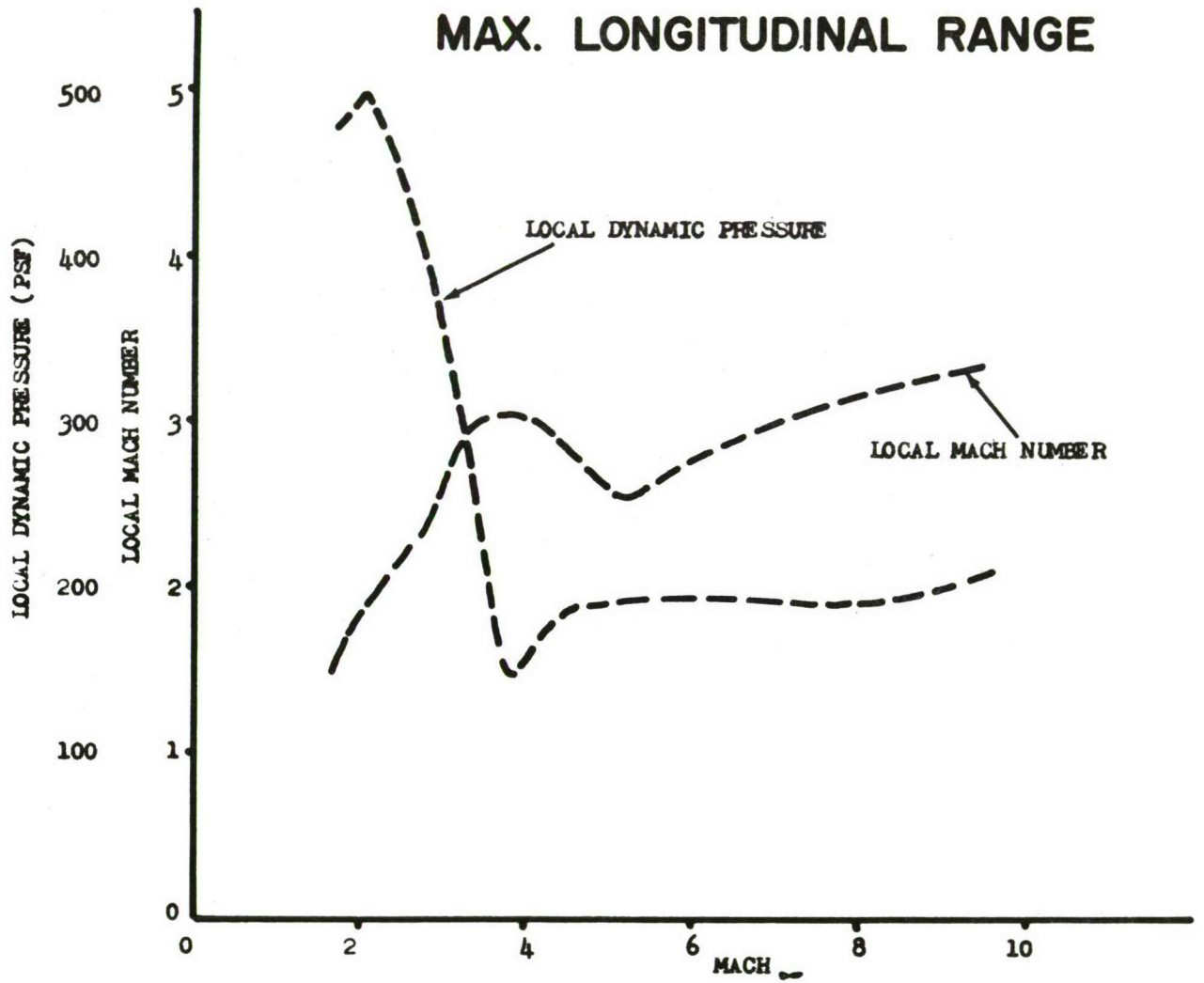


Fig. 14 Free Stream Mach No. vs Local Flow Conditions (Veh. #5)

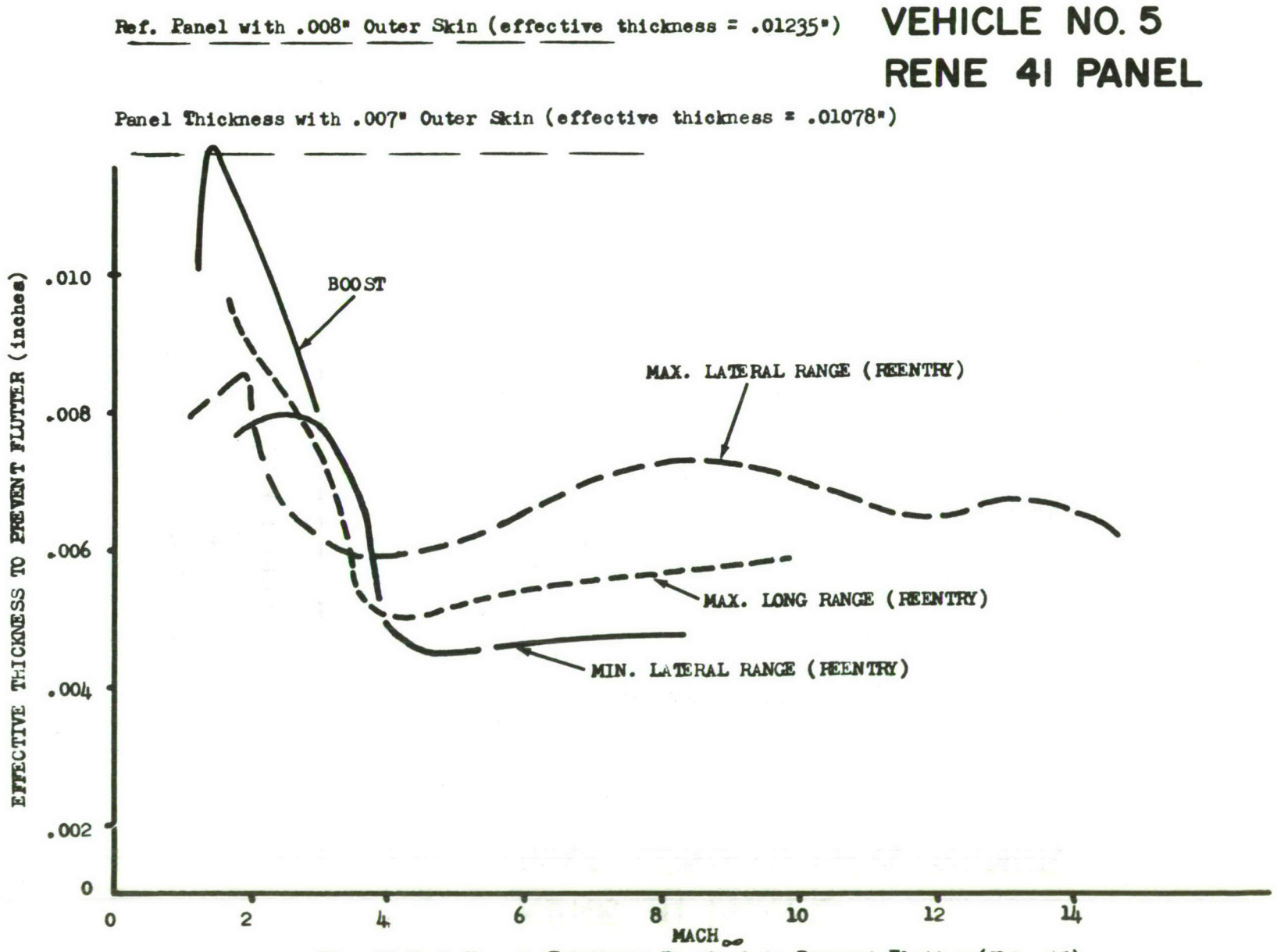


Fig. 15 Mach No. vs Thickness Required to Prevent Flutter (Veh. #5)

VEHICLE NO. 6
MOLYBDENUM PANEL

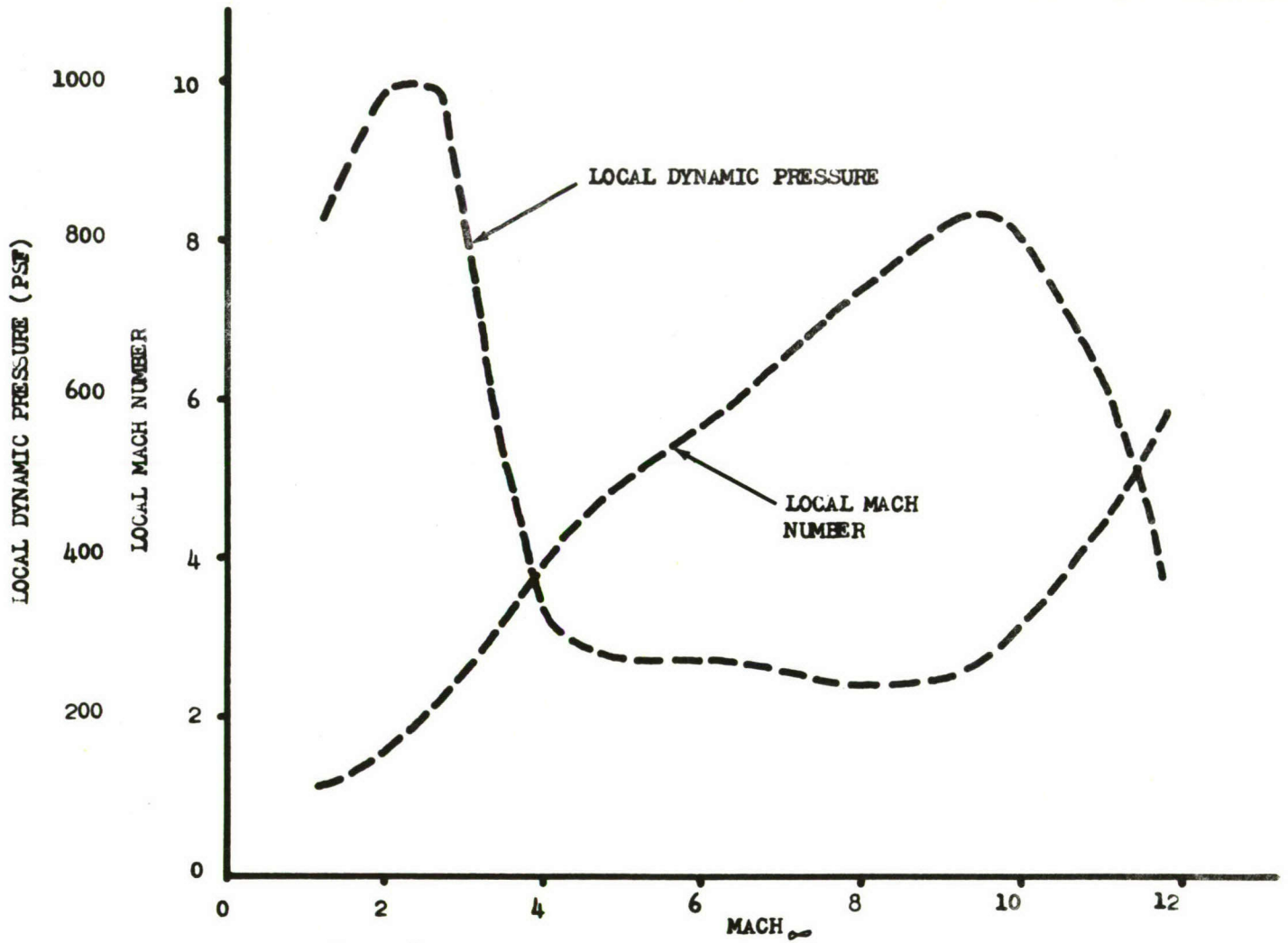


Fig. 16 Free Stream Mach No. vs Local Flow Conditions (Veh. #6)

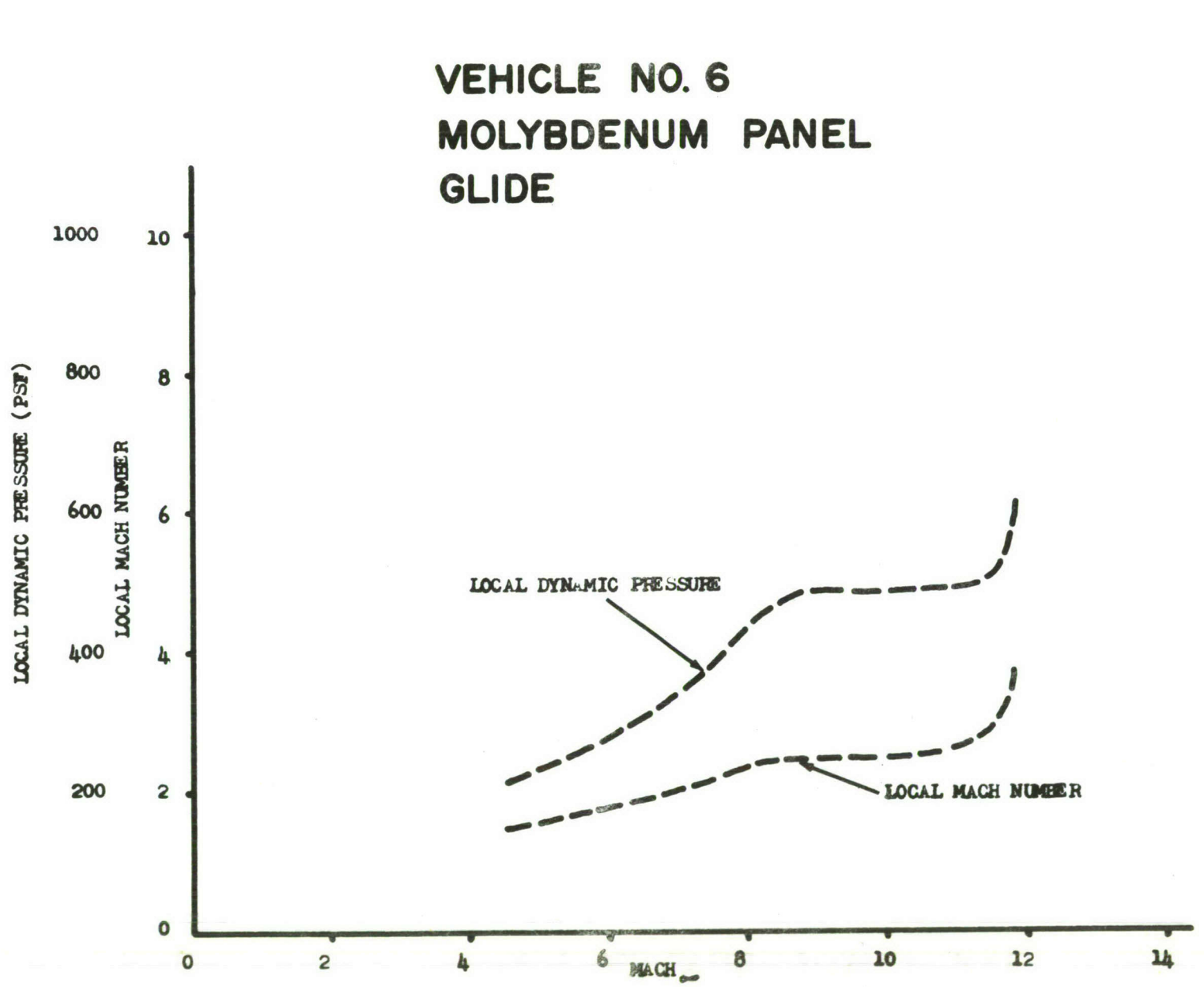


Fig. 17 Free Stream Mach No. vs Local Flow Conditions (Veh. #6)

VEHICLE NO. 6
MOLYBDENUM PANEL

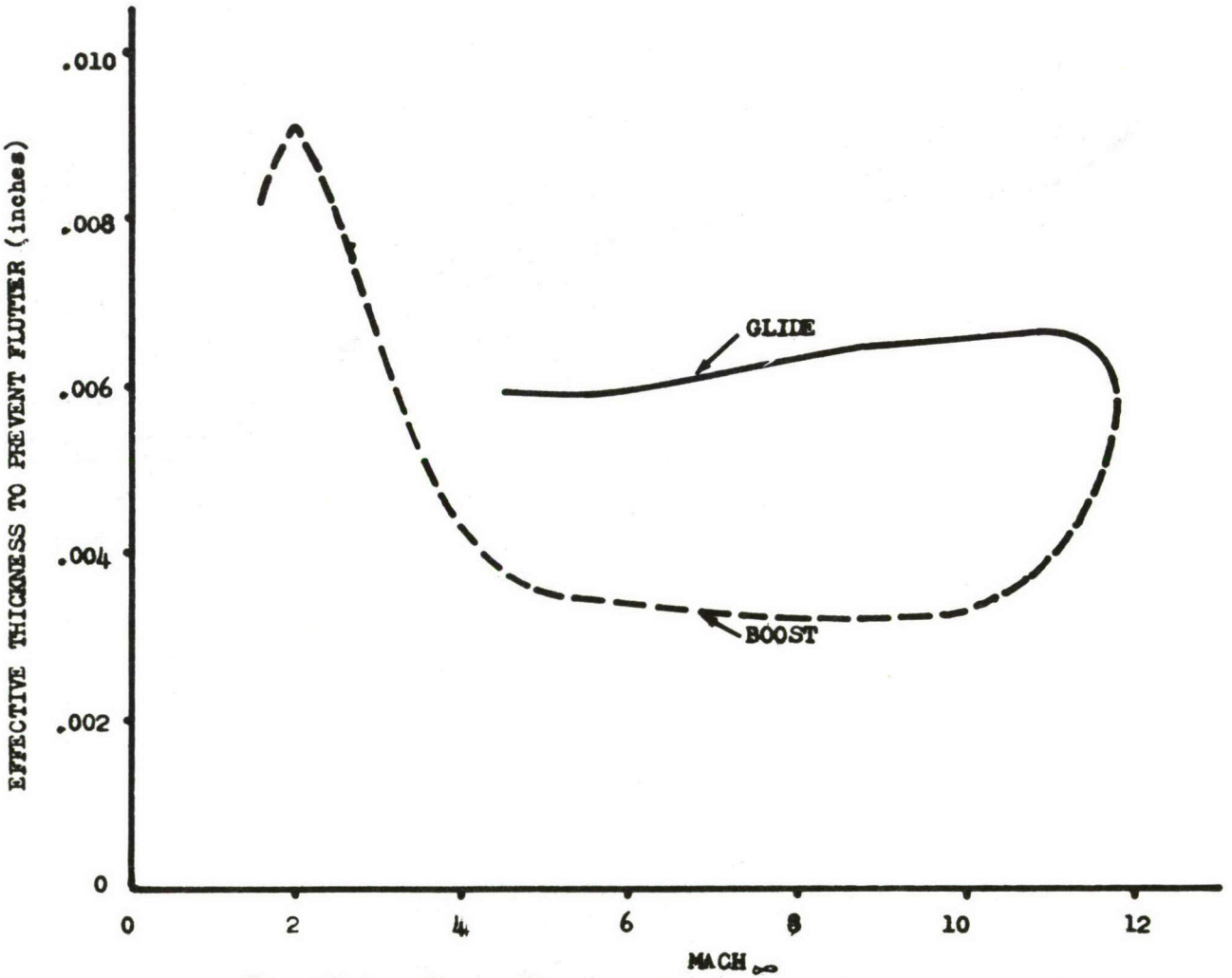


Fig. 18 Mach No. vs Thickness Required to Prevent Flutter (Veh. #6)

VEHICLE NO. 7
COLUMBIUM ALLOY PANEL

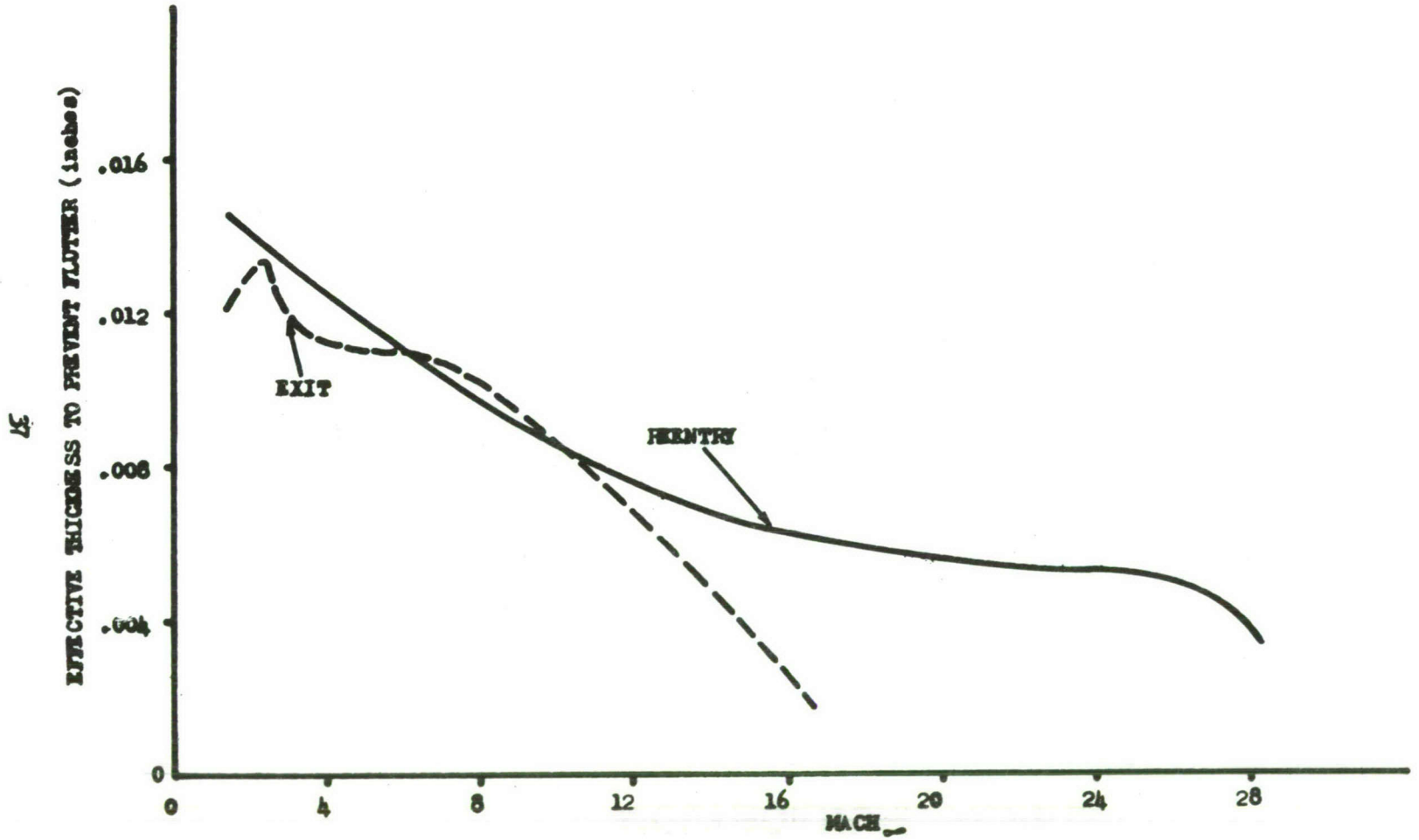


Fig. 19 Mach No. vs Thickness to Prevent Flutter (Veh. #7)

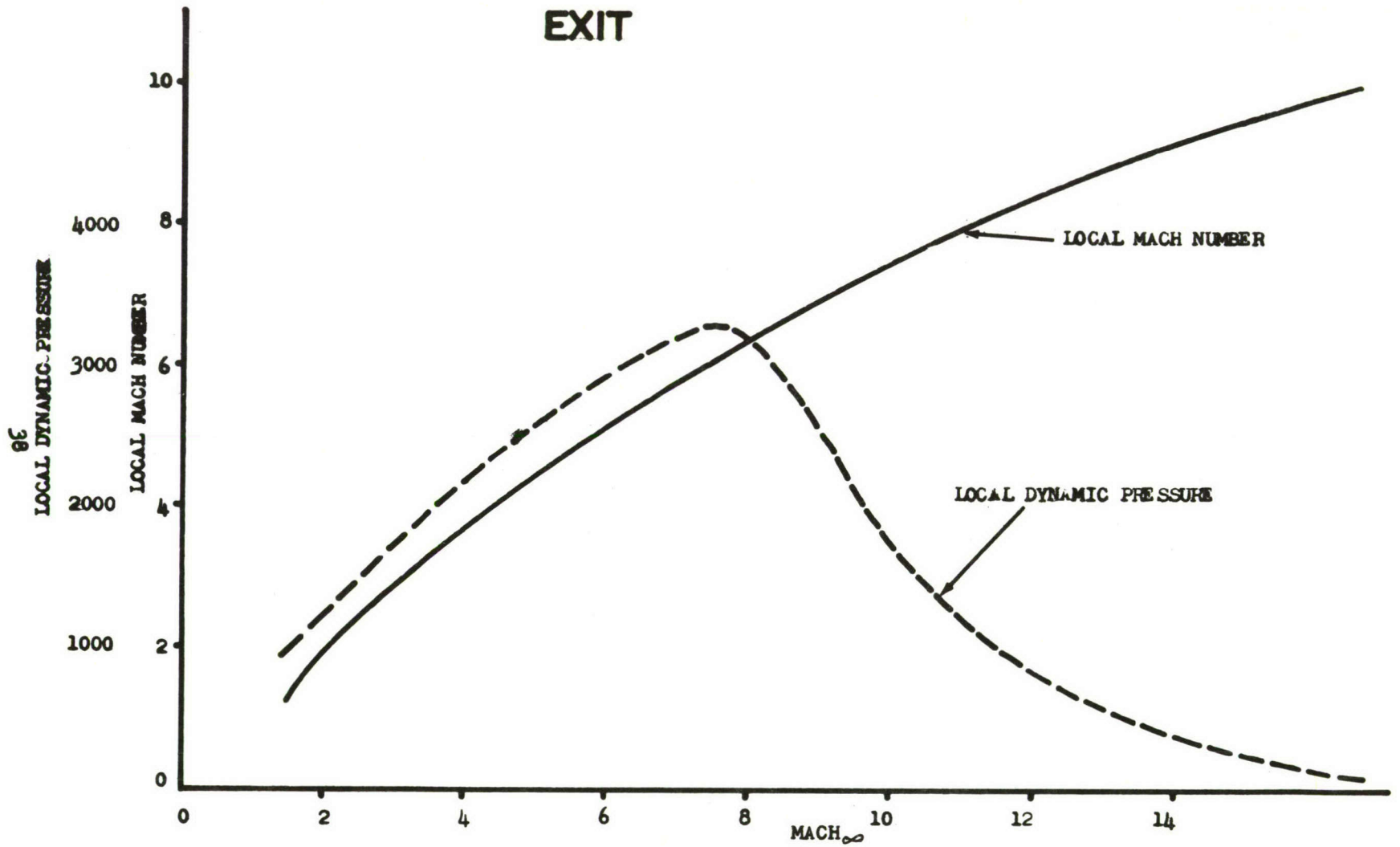


Fig. 20 Free Stream Mach No. vs Local Flow Conditioning (Veh. #7)

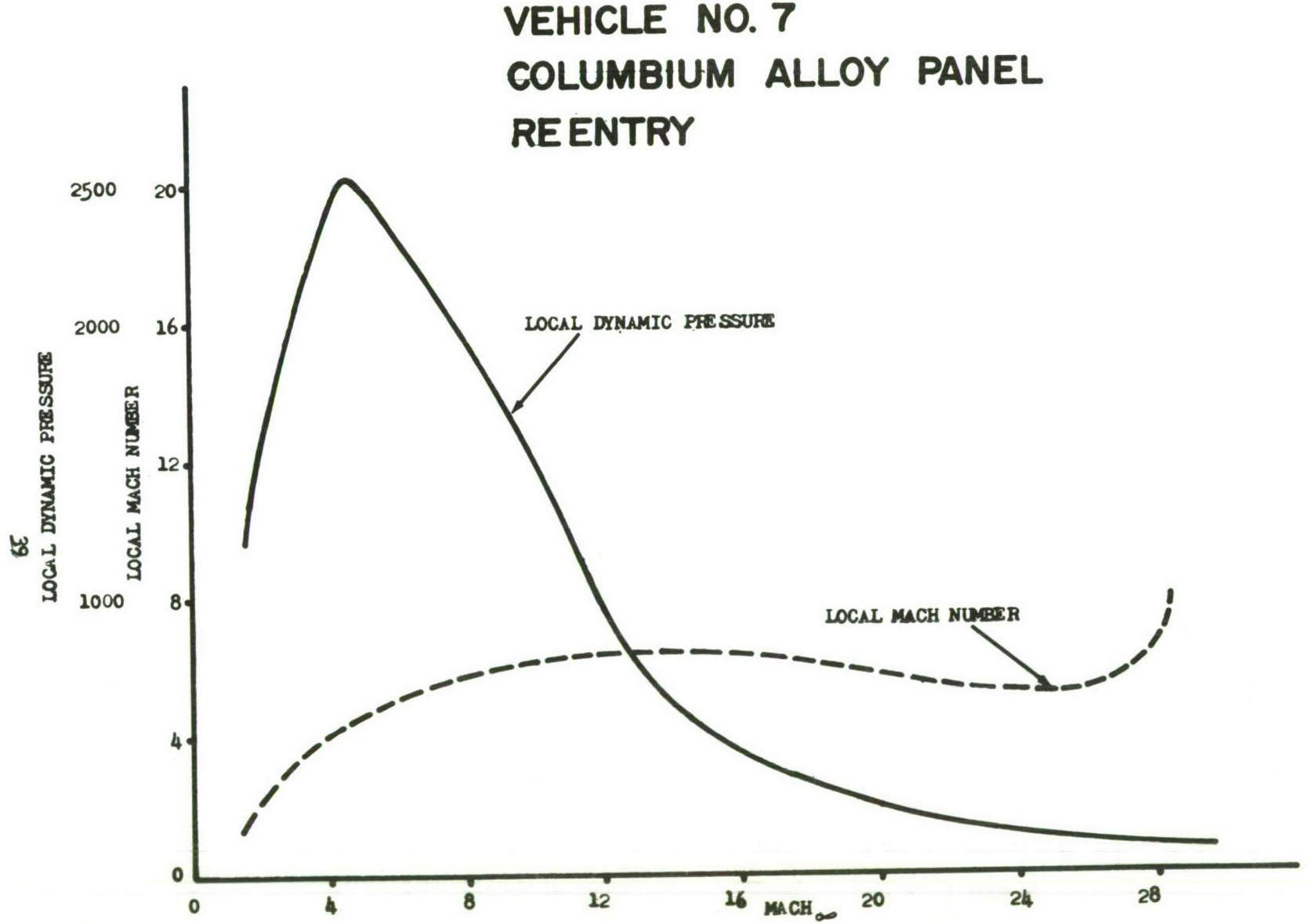


Fig. 21 Free Stream Mach No. vs Local Flow Conditioning (Veh. #7)

TABLE 1

SUMMARY OF COMPUTED PARAMETERS AND RATIOS

Vehicle		Max. Thickness to Prevent Flutter Conditions					Max. Local Dynamic Pressure Conditions			
		t_{mzx} (in.)	M	q_L (psf)	q_L/q_2	E_2/E_1	$q_{L\,mzx}$ (psf)	M	q_L/q_2	E_2/E_1
#1	Exit	.0102	4.15	3850	.68	.85	4230	7.2	.395	1.3
	Re-entry	.0054	9.9 (7.4)	860	.085	1.02	860	9.9	.085	1.02
#2	Exit	.0104	1.8	1750	.82	.75	2700	8.5	.225	.75
	Re-entry		Not Available							
#3	Exit	.0104	2.0	2150	.83	.71	2600	5.15	.39	.78
	Re-entry	.0063	11.8 (4.0)	650	.097	1.4	742	9.2	.108	1.0
#4	Exit	.0097	2.05	1600	.7	.7	2400	5.2	.35	.80
	Re-entry		Same as #1							
#5	Exit	.0109	1.5	1000	.72	1.0	1100	2.0	.49	1.0
	Re-entry	.0086	1.65 (1.65)	473	.286	1.2	500	2.1	.22	1.2
#6	Exit	.009	2.0	1000	.59	.68	1000	2.0	.59	.68
	Re-entry	.0065	11.4 (4.5)	486	.135	1.13	486	11.4	.135	1.13
#7	Exit	.013	2.3	1920	.66	1.95	3320	7.7	.34	1.95
	Re-entry	.0146	1.5 (1.5)	1220	.87	1.95	2560	4.5	.38	1.95

( ) Min. M available for re-entry

TABLE 2

MAXIMUM TEMPERATURES

	Exit		Re-entry	
Vehicle	Max. Temp. (°F)	M	Max. Temp. (°F)	M
#1	1650	24.5	2300	19.2
#2	1400	18.0	*	
#3	1790	11.6	2450	19.4
#4	1180	7.0	2300	19.2
#5	----		1590	16.8
#6	1200	10.5	1980	11.4
#7	1600	9.7	2130	265

* Re-entry Trajectory not available

ANALYTICAL INVESTIGATION OF EFFICIENCY IN STRUCTURAL VIBRATION WITH AND WITHOUT DAMPING TREATMENT

Channing L. Pao

JUNE 1964

Systems Engineering Group
Research and Technology Division
Directorate of Synthesis (SEST)
Deputy for Studies and Analyses
Wright-Patterson Air Force Base, Ohio

TABLE OF CONTENTS

	Page
LIST OF SYMBOLS	i
I. INTRODUCTION	1
1.1 General	1
1.2 Effect of Damping Treatments on Structural Vibration	2
1.3 Effect of Stiffness on Structural Vibration	3
1.4 Review of Previous Work	3
II. COMPARISON OF VIBRATIONAL ANALYSIS OF A SQUARE PANEL	
2.1 Objective	4
2.2 Analysis and Discussion of Damping Treatment Applicable to Harmonic Vibration	4
2.3 Analysis A Square Panel Subjected to Random Load	12
III. RESULTS	16
IV. CONCLUSIONS	17
BIBLIOGRAPHY	23

LIST OF SYMBOLS

a	length of side of square panel
B/A	stress ratio - stress without additional damping treatments ⋮ stress with additional damping treatments
D	structural damping factor before addition of damping treatment
E	Young's modulus of panel
$F_o e^{iwt}$	harmonic exciting force
$F(t)$	random exciting force
i	complex number
K	the flexural rigidity of the panel
k	$\frac{K\pi^4(m^2 + n^2)}{a^2}$ = stiffness of panel
m	mode number in (x) direction
n	mode number in (y) direction
Q_{mn}	generalized forces
q_{mn}	generalized displacement - solutions to Lagrange equation
q_o	displacement of panel at a particular point
rms	root mean square
T	total kinetic energy
$\overline{V}$	total strain energy
$W(x, y, t)$	amplitude with coordinates x and y and time t
$W_f(w)$	power spectral density of random exciting force
$W_g(w)$	power spectral density of generalized displacement
$\|Z\|^2$	square of modules of mechanical impedance
z	depth of panel

M	mass of panel
σ_x	flexural stress in (x) direction
σ_y	flexural stress in (y) direction
η	loss factor of damping treatment
δ	stress factor by which treatment increases the stiffness of the system
μ	mass factor by which treatment increase the mass of the system
ν	Poisson's ratio
w	circular forcing frequency
w_n	natural frequency

I. INTRODUCTION

1.1 General

With the advent of high powered aircraft and missiles, structural damage due to vibration and noise has become of increasing concern. Structures must withstand not only static and aerodynamic loads but also the destructive effects of high-peaked random loads. Conventional solutions such as rigidization, detuning and decoupling usually result in a weight increase.[1] The purpose of this report is to describe the initial exploratory work of optimum damping.

In general, mechanisms of damping in structures are:

1. Radiation or acoustic damping
2. Material damping
3. Interface, slip or joint damping
4. Viscoelastic and other special treatments

Pratical structural materials have internal hysteresis damping. Material hysteresis damping becomes less significant when interface and viscoelastic damping are considered.

Interface damping involves the relative motion of joints and fastenings of structural elements. Sliding friction between elements provides effective damping; however, its effectiveness varies with joint pressure and wear.

Viscoelastic damping is frequency and amplitude dependent; for the treatment to be effective at low frequencies, the dissipation per cycle must be increased, either by increasing the flexure per cycle or by using a material with greater inherent damping.

For maximum damping it is necessary to subject the viscoelastic layer to a high effective or over-all stress (or strain). High local shear stresses with low over-all stress do not produce high energy dissipation; the average interface viscoelastic stress must be high.

1.2 Effect of Damping Treatments on Structural Vibration

Large resonant amplification in conventional (low damped) structures creates a high resonant stress and shortens fatigue life. Adequate damping limits resonant amplitudes, and therefore reduces the peak stresses and extends the fatigue life of structural elements.

Viscoelastic damping in a built-up beam gives a large reduction in the resonant amplification for the first mode, and even greater reductions for higher modes, in comparison to a solid beam.[1] Tests have shown that interface damping of riveted joints in comparison to a solid beam can significantly reduce the response.[2]

Viscoelastic damping generally gives resonant amplification reductions which are greater than those offered by riveted joints. However, most viscoelastic materials are limited for extreme temperature variations. These materials lose much of their effectiveness at high or low temperatures (300° to -100° F).[1] Interface damping appears more attractive for such applications. A structural design that is highly damped is usually heavier than a conventional design, based on equal static stiffness. At frequencies far removed from resonance the system response is not greatly affected by damping.[3]

Damping can be undesirable if it causes dynamic instability of rotating shafts, if it reduces the accuracy of sensitive instruments and when internal heating is to be avoided.

1.3 Effect of Stiffness on Structural Vibration

The addition of structural stiffness to a panel increases its resonant frequency, lowers the vibration amplitude, and adds damping due to the material hysteresis of the stiffener and the associated joints and fasteners. Changing structural stiffness may change the natural frequency to either a higher or lower db level in the noise spectrum.[4] Stiffeners also provide a convenient means of adding viscoelastic material damping (at the interface).

1.4 Review of Previous Work

Early work was primarily concerned with the energy dissipation due to acoustic radiation and hysteretic effects in structural members. In some instances hysteresis damping has been sufficient to limit resonant stresses to reasonable levels. As the acoustic radiating power of jet and rocket engines increases, material damping becomes insufficient, and other mechanisms must be considered.

Initial studies on interface slip damping showed that large energy dissipation is possible under optimum pressure and geometry conditions. Optimum pressure for maximum slip damping can lead to fretting and corrosion of the interface surfaces, which in itself can be more serious than the sonic fatigue problem. Also, a small change in optimum conditions during service can cause a large reduction in damping and in the increase in resonant stress.*

*Studies initiated in 1955 showed that viscoelastic materials can dissipate considerably more energy than structural materials, per unit volume. See Reference (4).

New materials are being developed for broader applications. Mechanisms of joint and viscoelastic damping are the object of intensive research. Describing random loading involves complex mathematics and the results are often limited.

II. COMPARISON OF VIBRATIONAL ANALYSIS OF A SQUARE PANEL

2.1 Objective

The objective of this study is to analyze the vibrational characteristics of structural panels. A comparison is made between conventional panels and panels with additional damping treatments. The desired outcome is to determine which vibration control technique offers the greater potential, and whether to concentrate on material or design improvements.

A comparison is made of the vibration characteristics of simply supported symmetric square panels with harmonic excitation and with random excitation.

2.2 Analysis and Discussion of Damping Treatment Applicable To Harmonic Vibration

The analysis of the panels is based on the theory of normal modes, i.e., vibrational motion in which each particle of the structure vibrates in simple harmonic motion with the same frequency and passes through its equilibrium position at the same time. Total damping is determined by summing up the damping from each mode. Let a simply supported square panel be subjected to harmonic vibration as shown in Figure (1) and assume the first mode shape as,

$$W(x, y, t) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} q_{mn}(t) \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{a} \qquad (1)$$

where $w(x, y, t)$ is the amplitude of a vibrating panel with length and width equal to a. q_{mn} are solutions to the Lagrange Equation. m and n are the mode numbers in the x and y directions.

The total strain energy is

$$\overline{V} = \frac{K}{2}\int\int\left\{\left(\frac{\partial^2 W}{\partial x^2}\right)^2+\left(\frac{\partial^2 W}{\partial y^2}\right)^2+2\nu\frac{\partial^2 W}{\partial x^2}\frac{\partial^2 W}{\partial y^2}+2(1-\nu)\left(\frac{\partial^2 W}{\partial x\partial y}\right)^2\right\}dxdy \tag{2}$$

K is the flexural rigidity of the panel and is a function of EI and ν, where ν is Poisson's ratio

The total kinetic energy is

$$T = \frac{M}{2}\iint \dot{W}^2\, dxdy \tag{3}$$

where M = mass of panel

Equation (1) must satisfy the following boundary conditions:

at x = 0 and x = a; W = 0

at y = 0 and y = a; W = 0

$$\frac{\partial^2 W}{\partial x^2} = \frac{\partial^2 W}{\partial y^2} = 0$$

The displacements and moments at edges are zero,

Substituting Equation (1) into Equation (2) with the above boundary conditions yields

$$\overline{V} = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\frac{K q_{mn}^2\,\pi^4}{8\,a^2}\,(m^2+n^2)^2 \tag{4}$$

Substituting Equation (1) into Equation (3) gives

$$T = \frac{Ma^2}{8} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \dot{q}_{mn}^2 \qquad (5)$$

The Lagrange Equation is

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_{mn}} \right) - \frac{\partial T}{\partial q_{mn}} + \frac{\partial \bar{V}}{\partial q_{mn}} + \frac{\partial D}{\partial \dot{q}_{mn}} = Q_{mn} \qquad (6)$$

where D is the initial damping coefficient

and Q_{mn} are the generalized forces $= F_o e^{iwt}$

Substituting Equation (4) and Equation (5) into Equation (6) results in the equation of motion of the panel

$$\ddot{q}_{mn} + \frac{D}{M} \dot{q}_{mn} + \frac{K\pi^4}{M} \left(\frac{m^2+n^2}{a^2} \right) q_{mn} = \frac{Q_{mn}}{M} \qquad (7)$$

$$\frac{K\pi^4}{M} \left(\frac{m^2+n^2}{a^2} \right) = w_n^2$$

Let $k = K\pi^4 \left(\frac{m^2+n^2}{a^2} \right)$

Equation (7) becomes

$$Mq_{mn} + Dq_{mn} + kq_{mn} = F_o e^{iwt} \qquad (8)$$

When damping is added, the mass and stiffness increase from M and k to $M\mu$ and $k\delta(1 + i\eta)$, respectively. μ and η are factors which increase the mass and stiffness of the system and η is the loss factor of the damping treatment. The imaginary part of the complex stiffness is contributed by the loss factor of the added damping treatment.

The equation of motion of the panel with added damping becomes

$$M\mu\ddot{q}_{mn} + D\dot{q}_{mn} + K\delta(1 + i\eta)\, q_{mn} = F_o e^{iwt} \qquad (9)$$

The maximum value of q_{mn} occurs at the frequency

$$w_n = \sqrt{\frac{k\delta}{M\mu}} = \sqrt{\frac{\pi^4 K\left(\frac{m^2+n^2}{a^2}\right)}{M\mu}} \qquad (10)$$

Assume the solution of Equation (8) to be

$$q_{mn} = q_o e^{iwt}$$

Then, $\dot{q}_{mn} = q_o iwe^{iwt}$

$$\ddot{q}_{mn} = -q_o w^2 e^{iwt}$$

Substituting into Equation (8) gives

$$-Mq_o w^2 e^{iwt} + Dq_o iwe^{iwt} + kq_o e^{iwt} = F_o e^{iwt}$$

$$q_o = \frac{F_o}{-Mw^2 + Diw + k}$$

or

$$q_{mn} = \frac{F_o e^{iwt}}{Mw^2 + Diw + k} \qquad (11)$$

Substituting Equation (11) into Equation (1) gives the amplitude of vibration of the panel without additional damping.

$$w(x, y, t) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{F_o e^{iwt}}{-Mw^2 + Diw + k} \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{a} \qquad (12)$$

The solution for Equation (9) is assumed to be

$$q_{mn} = q_o e^{iwt}$$

Then, $\dot{q}_{mn} = q_o iwe^{iwt}$

$$\ddot{q}_{mn} = -q_o w^2 e^{iwt}$$

$$-M\mathcal{M} q_o w^2 e^{iwt} + Dq_o iwe^{iwt} + k\delta(1+i\eta)\, q_o e^{iwt} = F_o e^{iwt}$$

$$q_{mn} = \frac{F_o e^{iwt}}{-M\mathcal{M}w^2 + Diw + k\delta(1 + i\eta)} \qquad (13)$$

or

$$q_o = \frac{F_o}{-M\mathcal{M}w^2 + Diw + k\delta(1 + i\eta)}$$

Substituting Equation (13) into Equation (1) yields the amplitude of vibration of the panel with additional damping

$$w(x, y, t) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{F_o e^{iwt}}{-M\mathcal{M}w^2 + Diw + k\delta(1 + i\eta)} \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{a} \qquad (14)$$

The bending stresses of the panel in the x and y direction are: -

$$\sigma_x = \frac{E\,Z}{1-\nu^2} \left(\frac{\partial^2 W}{\partial x^2} + \nu \frac{\partial^2 W}{\partial y^2} \right)$$

$$\sigma_y = \frac{EZ}{1-\nu^2} \left(\frac{\partial^2 W}{\partial y^2} + \nu \frac{\partial^2 W}{\partial x^2} \right) \qquad (15)$$

Since the panel is assumed to be symmetric, σ_x and σ_y are therefore equal.

Substituting Equation (12) into Equation (15) gives

$$\sigma_x = \sigma_y = \frac{-E\,z\;F_o e^{iwt}}{(1-\nu^2)(Diw+k-Mw^2)} \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \left(\frac{m^2\pi^2 + n^2\pi^2\nu}{a^2}\right) \sin\frac{m\pi x}{a} \sin\frac{n\pi y}{a} \tag{16}$$

This equation gives the bending stress of the initial damped panel; similarly substituting Equation (14) into Equation (15) yields the bending stress of the panel with additional damping treatment.

$$\sigma_x = \sigma_y = \left(\frac{-E\,z}{1-\nu^2}\right) \frac{F_o e^{iwt}}{Diw+k\delta(1+i\eta) - M\mu w^2} \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \left(\frac{m^2\pi^2+n^2\pi^2\nu}{a^2}\right) \sin\frac{m\pi x}{a} \sin\frac{n\pi y}{a} \tag{17}$$

Let A = Bending stress in panel with additional damping treatments. (Equation 17)

B = Bending stress in panel without additional damping treatments. (Equation 16)

Equation (16) divided by Equation (17) is

$$\frac{B}{A} = \frac{Diw + k\,\delta(1+i\,\eta) - M\mu w^2}{Diw + k - Mw^2} \tag{18}$$

$$\text{or } i\,(ADw+Ak\delta\eta - BDw) + Ak\delta - AM\mu w^2 - Bk + BMw^2 = 0 \tag{19}$$

In order to satisfy Equation (19), the complex and real terms must each be equal to zero.

$$i\,(ADw + Ak\delta\eta - BDw) = 0 \tag{20}$$

and

$$Ak\delta - AM\mu w^2 - Bk + BMw^2 = 0 \tag{21}$$

Equation (20), divided by A, and with terms rearranged is

$$\frac{B}{A} = 1 + \frac{k\delta\eta}{Dw}, \quad \frac{A}{B} = \frac{Dw}{Dw + k\delta\eta} \tag{22}$$

$$\text{or } \frac{A}{B} = \frac{1}{1 + \dfrac{k\delta\eta}{Dw}} = \frac{1}{1 + \left(\dfrac{w_n}{w}\right)\dfrac{(M\mu)^{1/2}(k\delta)^{1/2}\eta}{D}} \tag{23}$$

Equation (23) shows the stress ratio of a panel with added damping treatment to a panel without additional damping treatment as a function of loss factor, stiffness, mass factor and frequency ratio. In the damping parameter, stiffness is to the one half power, whereas the loss factor is to the first power. This indicates that an increase in the loss factor has a greater effect than an equal increase in stiffness. This equation expressed in the form of Equation (22) is plotted in Figure (2). It can be seen that the stress ratio is most influenced by an increase in the loss factor at the lower frequency range. Equation (21) divided by A, and with terms rearranged becomes

$$\frac{B}{A} = \frac{M\mu w^2 + k\delta}{Mw^2 - k}$$

or

$$\frac{B}{A} = \frac{\mu - \dfrac{k\delta}{Mw^2}}{1 - \dfrac{k}{Mw^2}} \tag{24}$$

Substituting Equation (23) into Equation (24) gives

$$\mu - \frac{k\delta}{Mw^2} = \left(1 + \frac{k\delta\eta}{Dw}\right)\left(1 - \frac{k}{Mw^2}\right) \qquad (25)$$

Substituting $w_n^2 = \frac{k\delta}{M\eta}$ into the above equation and rearranging terms gives

$$\mu = 1 + \frac{(k\delta M\mu)^{1/2}\eta}{D}\left(\frac{w}{w_n}\right)\left[1 - \frac{\mu}{\delta}\left(\frac{w_n}{w}\right)^2\right] + \left(1 - \frac{1}{\delta}\right)\left(\frac{w_n}{w}\right)^2\mu$$

The last term of the above equation is small in comparison with the other terms. Ignoring this term yields

$$\frac{(k\delta M)^{1/2}}{D}\eta = \left(\frac{w}{w_n}\right)^3\left[\frac{\mu - 1}{\left(\frac{w}{w_n}\right)^2 - 1}\right]\frac{1}{(\mu)^{1/2}} \qquad (26)$$

Equation (26) shows the effect of damping treatment mass and frequency ratio on loss factor. At a given frequency ratio, small increases in mass of damping treatment produce large changes in loss factor. For instance, at frequency ratio 5, increasing the mass factor from 1.1 to 1.25 more than doubles the loss factor. This equation is plotted in Figure (3). Note that all the mass ratios (or mass increase factors) and their tangential points are located on the same frequency ratio axis. The tangential point is the lowest loss factor point. Smaller loss factor will result in smaller damping in a particular vibrating system. This also indicates that for all cases of mass increase factor, μ, the minimum damping effect by increasing mass factors is at the frequency ratio of 1.15. It appears that the effectiveness of additonal damping treatment causing increased mass factors is more effective in the higher

frequency ratio region.

Combining Equations (26) and (23) yields the following relationship:

$$\frac{B}{A} = 1 - \left(\frac{w}{w_n}\right)^2 \left[\frac{\mu - 1}{\left(\frac{w}{w_n}\right) - 1}\right] \qquad (27)$$

Equation (27) indicates that stress ratio is shown to be a function of damping mass and frequency ratio. At any given frequency ratio, stress ratio can be found for any given mass factor of the treatment. The stress ratio is frequency dependent, and has its maximum values at resonance. This equation is plotted in Figure (4) and is discussed further in Section III, RESULTS.

2.3 Analysis A Square Panel Subjected to Random Load.

Let the system be excited by a random force. Assume that the power spectrum of the corresponding generalized force does not vary appreciably in the region of the natural frequency of the system.

The equation of motion with a random exciting force becomes

$$M\mu \ddot{q}_{mn} + D\dot{q}_{mn} + k\delta(1+i\eta)\, q_{mn} = F(t) \qquad (28)$$

If the power spectral density of $F(t)$ is denoted by $W_f(w)$, then the power spectral density of the generalized displacement $W_g(w)$ is given by

$$Wg(w) = \frac{W_f(w)}{|Z|^2} \qquad (29)$$

where $|Z|^2$ is the square of the mechanical impedance, which is

$$(k\delta - w^2 M\mu)^2 + (wD + k\delta\eta)^2$$

The mean square value of the generalized displacement for the panel with added damping is

$$\overline{q}^2 = \int_0^\infty W g(w)\,dw$$

$$= \int_0^\infty \frac{W_f(w)}{|Z|^2}\,dw = \frac{W_f(w_n)\frac{\pi}{2}}{k\delta\left(D + \frac{k\delta\eta}{w_n}\right)} \qquad (30)$$

Where $w_n^2 = \frac{k\delta}{M\mu}$

The root mean square (rms) amplitude is

$$q_{rms} = \left[W_f(w_n)\,\frac{\pi}{2}\right]^{1/2} \left[k\delta\left(D + \frac{k\delta\eta}{w_n}\right)\right]^{-1/2} \qquad (31)$$

Similarly, without added damping, the rms amplitude is

$$q_{rms} = \left[W_f(w_n)\,\frac{\pi}{2}\right]^{1/2} [kD]^{-1/2} \qquad (32)$$

Substituting Equation (31) into Equation (1) gives the amplitude with additional damping

$$W(x, y, t) = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \left[W_f(w_n)\,\frac{\pi}{2}\right]^{1/2} \left[k\delta\left(D + \frac{k\delta\eta}{w_n}\right)\right]^{-1/2} \sin\frac{m\pi x}{a} \sin\frac{n\pi y}{a} \qquad (33)$$

Substituting Equation (32) into Equation (1) gives the amplitude without additional damping

$$W(x, y, t) = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \left[W_f(w_n) \frac{\pi}{2}\right]^{1/2} [kD]^{-1/2} \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{a} \qquad (34)$$

Equation (15) is

$$\sigma_x = \sigma_y = \frac{Ez}{1-\nu^2} \left(\frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 W}{\partial y^2}\right)$$

Thus, the bending stress for the panel with additional damping is

$$\sigma_x = \sigma_y = \frac{Ez}{1-\nu^2} \left[W_f(w_n) \frac{\pi}{2}\right]^{1/2} \left[k\delta\left(D + \frac{k\delta\eta}{w_n}\right)\right]^{-1/2}$$

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \left(\frac{m^2\pi^2 + n^2\pi^2\nu}{a^2}\right) \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{a} \qquad (35)$$

Similarly, the bending stress for the panel without additional damping is

$$\sigma_x = \sigma_y = \frac{Ez}{1-\nu^2} \left[W_f(w_n) \frac{\pi}{2}\right]^{1/2} [kD]^{-1/2}$$

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \left(\frac{m^2\pi^2 + n^2\pi^2\nu}{a^2}\right) \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{a} \qquad (36)$$

Let A = Bending stress in panel with additional damping treatments (Equation 35)

B = Bending stress in panel without additional damping treatments (Equation 36)

Then, Equation (36) divided by Equation (35) is

$$\frac{B}{A} = \delta^{1/2} \left[1 + \frac{k \delta \eta}{D \, w_n} \right]^{1/2} \quad ; \quad w_n = \left(\frac{k \delta}{M \mu} \right)^{1/2}$$

$$\frac{B}{A} = \delta^{1/2} \left[1 + \frac{(k \delta M \mu)^{1/2}}{D}^{1/2} \eta \right] \tag{37}$$

δ is a function of μ. Assuming that δ is not appreciably greater than one, the square root of δ will then be even closer to one. Ignoring $(\delta)^{1/2}$, Equation (37) becomes

$$\frac{B}{A} = \left[1 + \frac{(k \delta M \mu)^{1/2}}{D} \eta \right]^{1/2} \tag{38}$$

Equation (38) is similar to Equation (23), except for its power (square root) and its independence of frequency ratio. Stress ratio is function of the square root of the loss factor and to the fourth root of the stiffness, stiffness and mass factors, and mass. Thus, loss factor increases have more influence on stress ratio than stiffness increases.

III. RESULTS

In Figure (2) are shown the effects of additional damping treatments on stress ratio A/B, (stress in a panel with damping treatments divided by stress in a panel without damping treatments). For example, when D/w = 0.1 and damping parameter $k\delta\eta$ = 0.005, A = 0.94B. When D/w = 0.1 and damping parameter $k\delta\eta$ = 0.5, A = 0.16B. Therefore it is evident that the stresses in the panel are reduced by additional damping. However, these curves indicate there is more reduction of stresses in the panel at the higher frequency region as compared with the lower frequency region.

In Figure (3) the effect of damping mass and frequency ratio on the loss factor is shown. At a given frequency ratio, a small increase in damping mass, μ, will produce a large change in the loss factor.

Figure (4) is a combination of Figures (2) and (3). The left side of the curve is plotted for stress ratios greater than one. The dotted curves indicate negative ratios which are due to a phase shift between the force and the motion at resonance.

From Figure 5, note that for values of the loss factor, η, near zero, the stress ratio rises steeply as the stiffness parameter increases. For a constant value of stiffness parameter, the stress ratio increases rapidly as η increases to about 0.2, and then flattens out. Therefore it may be concluded that in order to reduce the stress in Panel A, added damping is more effective if the loss factor is less than 0.2. Comparison of Figure 5 with Figure 4 shows that damping treatments under random excitation do not give as large a stress reduction as do treatments at resonant harmonic excitation.

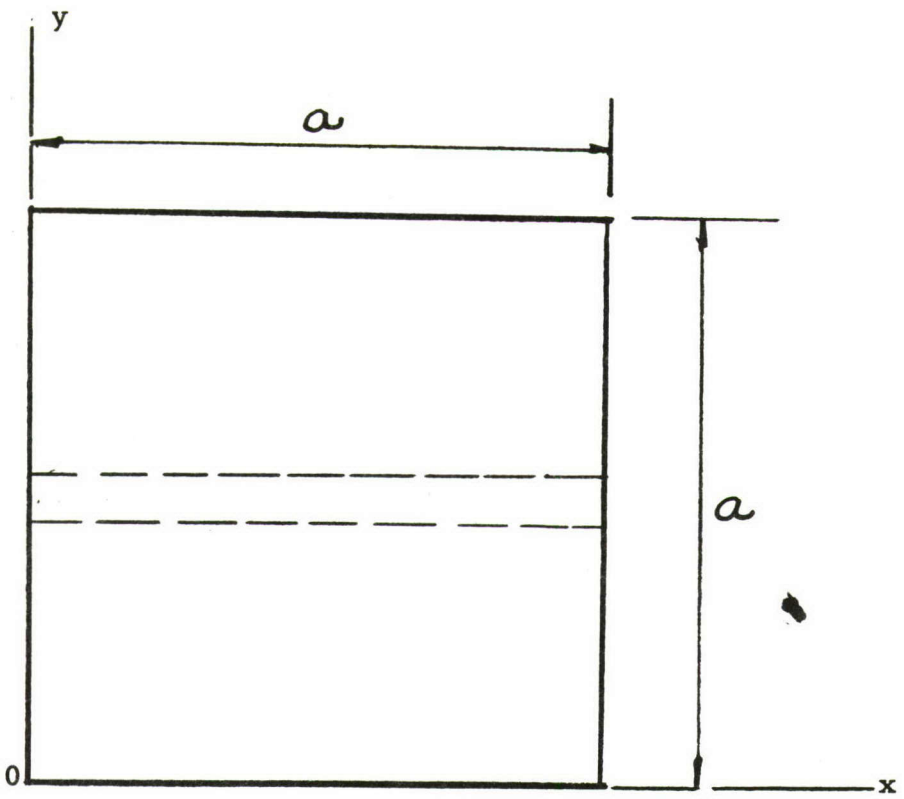
IV. CONCLUSIONS

Damping treatments are effective in reducing resonant stress amplitudes. A small amount of treatment can give a large resonant stress reduction. Good design in optimum damping is important to keep the weight at a minimum. Structures with optimum damping need not be designed for much more than static and aerodynamic loads. Increasing stiffness is effective when resonant vibrations can be avoided.

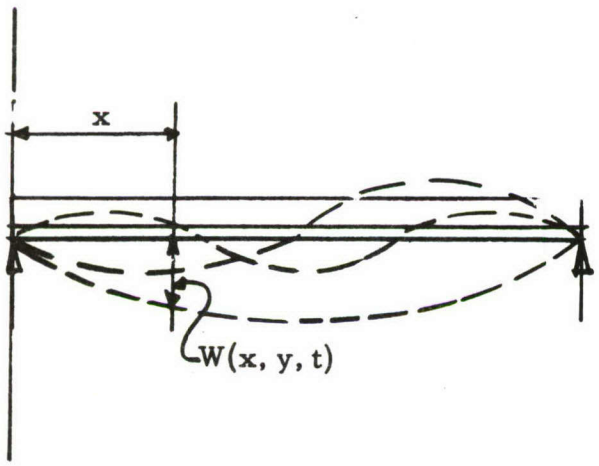
Under random loading, damping treatments are more beneficial than increasing stiffness. Increasing stiffness alone is of little value if resonance cannot be avoided.

Optimum damping in the harmonic vibrations differs from the optimum damping for random conditions. Effectiveness in suppressing vibration amplitudes cannot be sufficiently defined by the loss factor of the treatment.

There are few data in the literature on optimum damping. More experimental study is needed in this field.



SIMULATED AIRCRAFT PANEL



NORMAL MODES OF A PANEL BENDING

FIGURE 1 A SIMPLY SUPPORTED PANEL

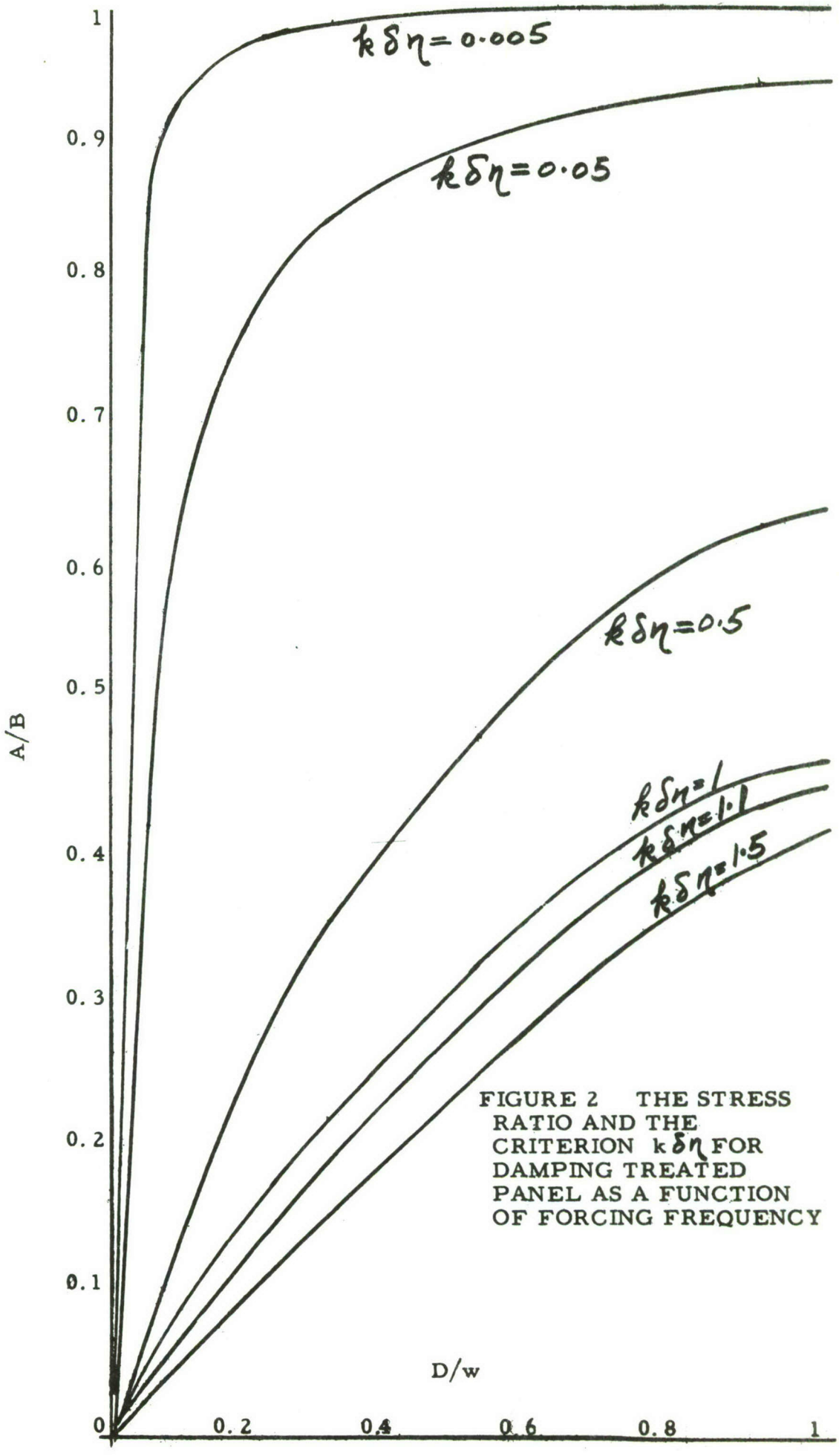


FIGURE 2 THE STRESS RATIO AND THE CRITERION $k\delta\eta$ FOR DAMPING TREATED PANEL AS A FUNCTION OF FORCING FREQUENCY

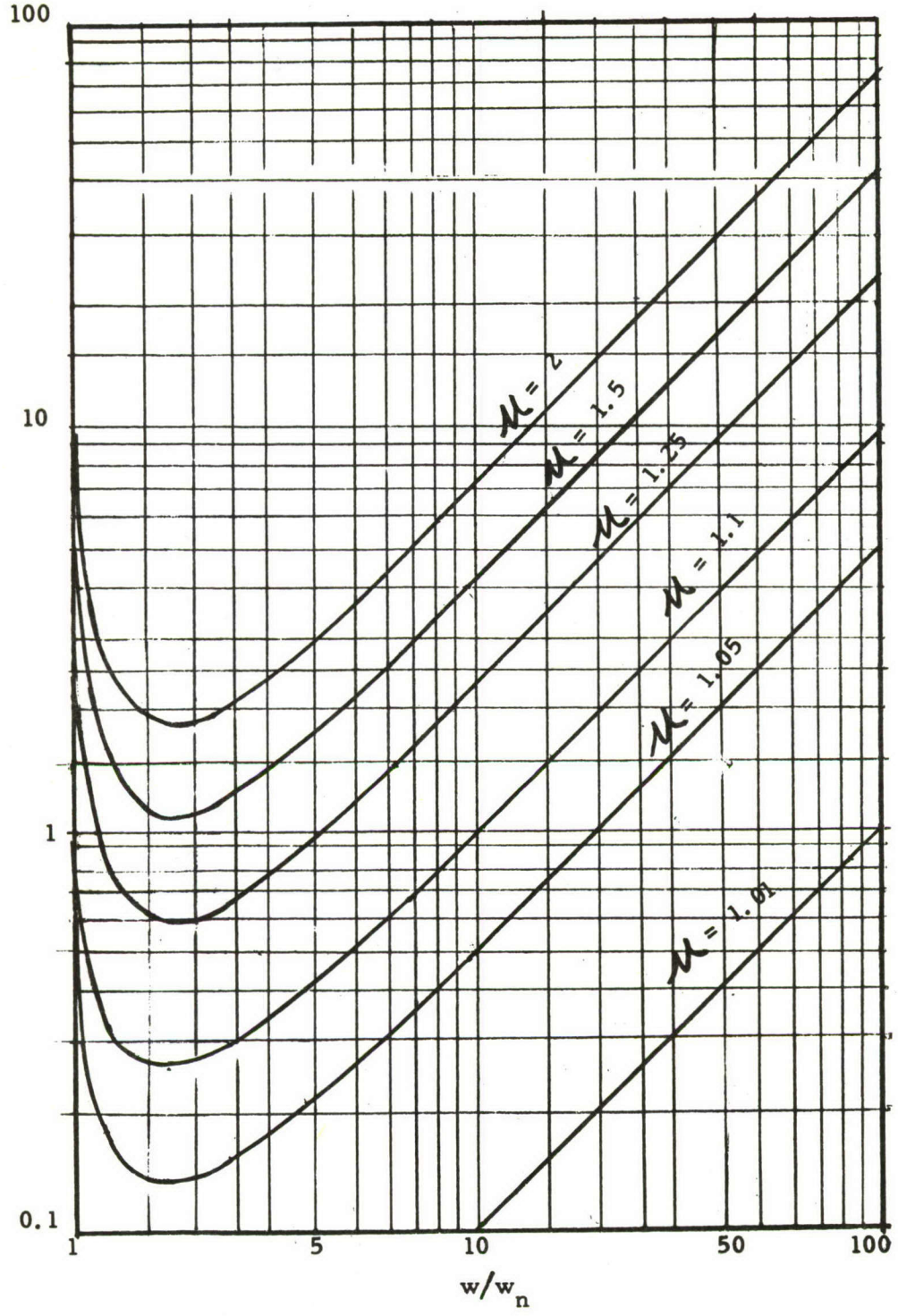
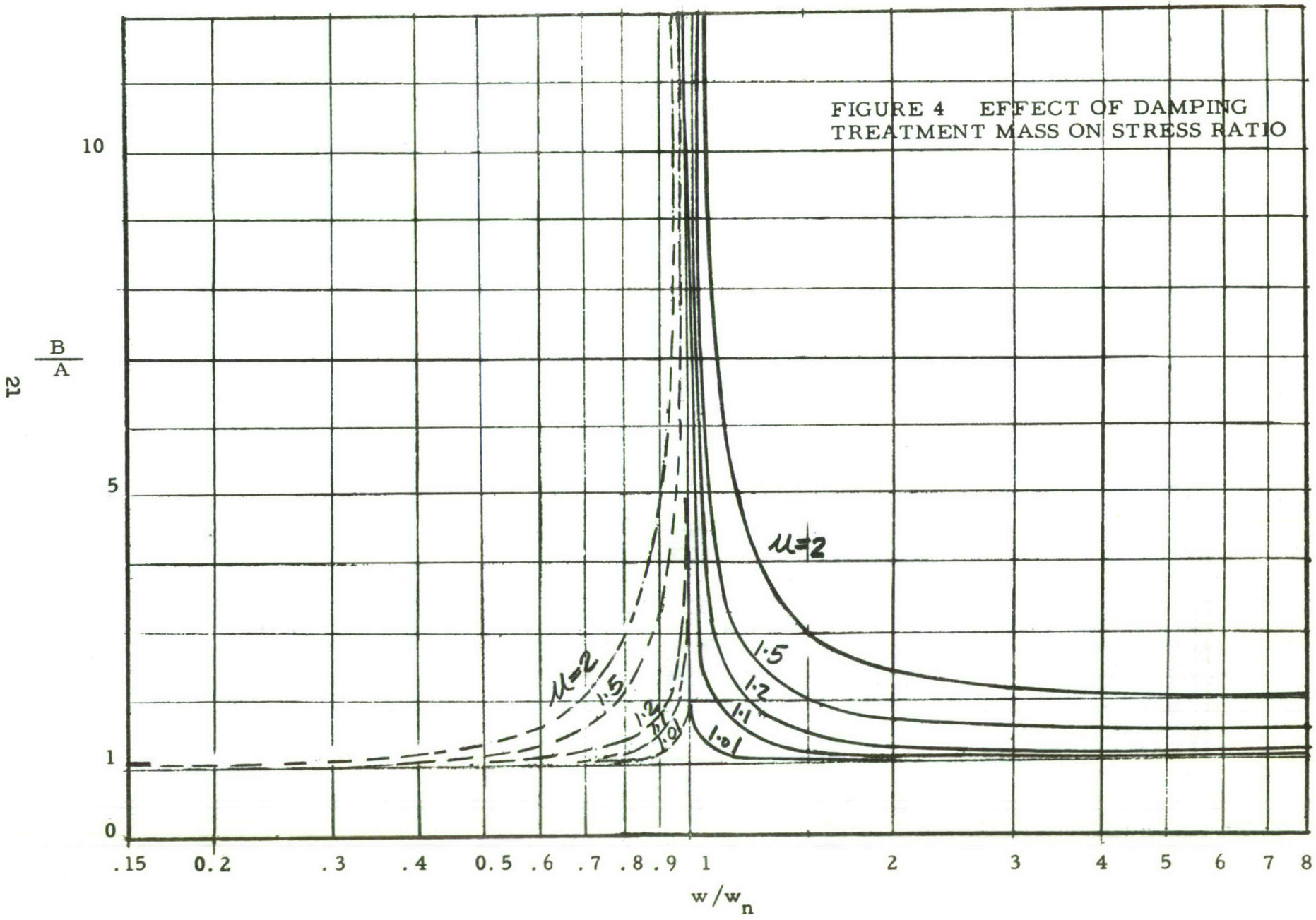


FIGURE 3 FREQUENCY RATIO AND LOSS FACTOR AS A FUNCTION OF RELATIVE WEIGHT OF TREATMENT


FIGURE 4 EFFECT OF DAMPING
TREATMENT MASS ON STRESS RATIO
B/A
10
5
1
0
.15
0.2
.3
.4
0.5
.6
.7
.8
.9
1
2
3
4
5
6
7
8
w/w_n
μ=2
1·5
1·2
1·1
1·01
μ=2
1·5
1·2
1·1
1·01

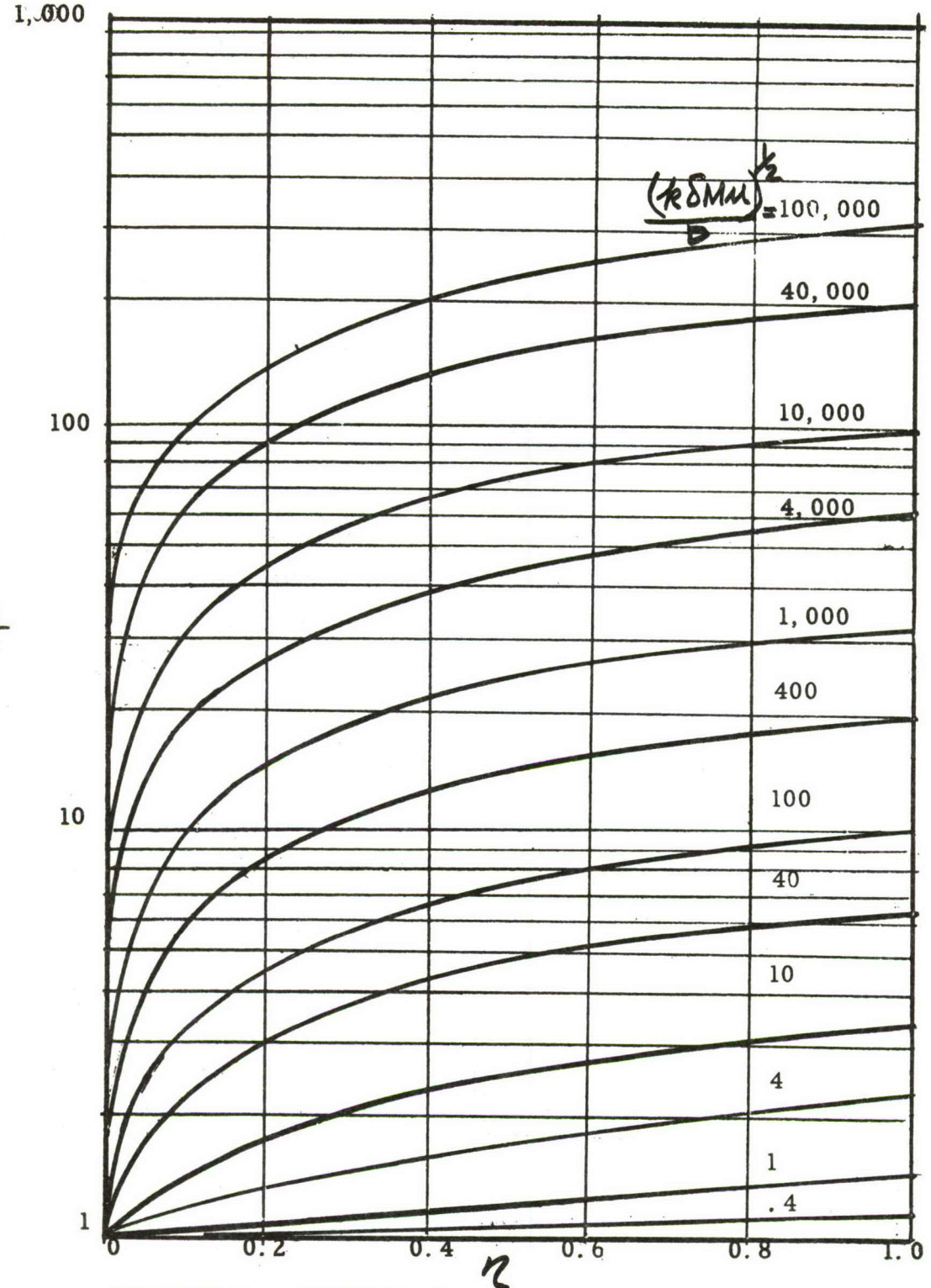


FIGURE 5 EFFECT OF ADDITIONAL DAMPING TREATMENTS ON STRESS RATIO RANDOM EXCITATION

BIBLIOGRAPHY

1. Jerome E. Rusicka, Staff Engineer, Berry Controls, Inc., "Builtin Damping: New Answer to Structural Vibration," Space/Aeronautics, April 1960, pp. 61-72.

2. Moses Tavvil, "Structural Damping Improved by Riveting," Space/Aeronautics, November 1962, pp. 97- 104.

3. Stephen H. Crandall, "Dynamic Response of Systems With Structural Damping," Air Force Office of Scientific Research Report AFOSR 1561, ASTIA AD No. 266299, October 1961, pp. 1-2.

4. W.J. Trapp, B.J. Lazan, "Role of Structural Damping in Acoustical Fatigue," WADC Technical Report 59-304, January 1960, pp. 6, 8, and 9.

5. Donald Ross, Eric E. Ungar, E.H. Kerwin, Jr., "Damping of Plate Flexural Vibrations by Means of Viscoelastic Laminae," Structural Damping, ASNE Annual Meeting, December 1959. Section 3, pp. 64, 68, and 83.

6. B.J. Lazan, "Energy Dissipation Mechnaisms in Structures With Particular Reference to Material Damping," Structural Damping, ASME Annual Meeting, December 1959, Section, p. 18.

7. Eric E. Ungar, Donald Ross, Edward M. Kerwin, Jr., "Damping of Flexural Vibrations by Alternate Viscoelastic and Elastic Layers," WADC Technical Report 59-509, November 1959, p.1.

8. I. Torbe, D.I.G. Jones, "On Influence Functions in the Theory of Forced Vibrations of Membranes," WADD Technical Report 61-97, March 1961.

9. G.G. Parfitt, D. Lambeth, "The Damping of Structural Vibrations," Aeronautic Research Council A.R.C. 22, 250, 0.1580, N.143, ASTIA AD No. 253216, September 1960.

10. Robert W. Hess, Leslie W. Lassiter and Harvey R. Hubbard, "A Study of the Responses of Panels to Random Acoustic Excitation," NACA RM L55E13c, July 15, 1955.

11. C.T. Heu, C.W. Chu, C.C. Chang, "Effect of Viscoelastic Foundation on Forced Vibration of Loaded Rectangular Plates," WADD Technical Report 60-360, July 1960.

12. T.J. Mentel, C.C. Fu, "Analytical Formulation of Damped Stress-Strain Relations Based on Experimental Data With Applications to Vibration Structures," ASD Technical Report 61-623.

13. N. Oberst, G. W. Becker and K. Frankenfield, "The Damping of Vibrations of Thin Sheet Metal by Means of Firmly Adhering Coatings," Acoustics (1954) 433-444, A.C.S.I.L. Translation No. 789, ASTIA AD No. 134731.

14. Stephan H. Crandall, Asim Yildiz, "Random Vibration of Beams," Air Force Office of Scientific Research AFOSR 484, ASTIA AD No. 256549, April 1961.

15. Robert F. Lambert, "Response of Bars (With Internal and Boundary Damping) to Transient and Random Excitation," WADC Technical Report 58-569, ASTIA Document No. 214382, May 1959.

16. Timoshenko, S. "Viberation Problems In Engineering." D. Van Nostrand Company, Inc.

17. Crandall, S.H., "Random Vibration," John Wiley and Sons, Inc.

(U) Crack Propagation in Cumulative Damage Fatigue Tests.

by Harry A. Lipsitt*

Abstract

Although considerable fatigue research has been aimed at predicting, measuring and attempting to understand crack propagation, such studies are meagre and of a cursory nature for cumulative damage fatigue. There are several existing theories in this area, none of them satisfactory. The simplest, called Miner's Law, is based on the assumption that the damage at each stress level is proportional to the number of cycles at that stress.

In this paper the failures in Miner's Law are discussed. Several specific reasons for this failure are mentioned and an experiment outlined to provide a quantitative assessment of the importance of each of these facets. A new theoretical approach to cumulative damage is presented and an experiment designed to unequivocally test the new hypothesis.

The data gathered for these experiments consists of measurements in single stress and cumulatively damaged fatigue specimens of crack density and crack length as a function of depth below the surface of the specimen, stress spectra, and fatigue cycles applied. These data are presented and discussed in terms of the previously described ideas and are shown to partially support the ideas developed in this paper. Further, the existence is shown of a band of previously unknown fatigue cracks originating within the body of the specimen, and the nature and significance of these cracks is discussed.

---

*Harry A. Lipsitt is Supervisor, Metallurgy Research, Metallurgy and Ceramics Research Laboratory, Aerospace Research Laboratories, Office of Aerospace Research, Wright-Patterson Air Force Base. Ohio.

INTRODUCTION

Structures in service under dynamic conditions are subject to stress cycles of rapidly varying amplitude, often applied in a more or less random sequence. It is easy to understand, then, that the results of conventional constant amplitude fatigue tests are not readily extrapolated to predict actual fatigue performance. This, however, has been the only thing an engineer could do when considering crack propagation in cumulative damage; all the extant data on crack propagation derive from single stress fatigue tests.

There are several existing theories in this area, none of which are wholly satisfactory (1-3). The simplest and most widely used, is called Miner's Law, or the linear damage law. This paper is a interim report of the results in a long term study of crack propagation in fatigue. It is specifically concerned with an attempt to rationalize the failures in Miner's Law. This concept holds that the existing damage in a fatigue specimen may be expressed in terms of the number of cycles applied divided by the number required to produce failure at a given stress level. If cycles of more than one stress are applied the cumulative damage is the sum of the individual "damage increments". Failure from cumulative damage is assumed to occur when the summation of the damage increments equals unity.

Mathematically

$$\sum \frac{n}{N} = 1$$

where n = number of cycles applied at stress S

N = number of cycles to failure at the same stress S

n/N = C, the damage increment, or more commonly, the cycle ratio.

The most obvious and puzzling failure of this wonderfully simple concept is that the cumulative cycle ratio for failure of a test piece may be changed (from less than unity to greater than unity) merely by reversing the sequence in which two different levels of alternating stress are applied. For example, it has been shown[4] that when a specimen is stressed first to 60% cycle ratio at a stress S_1 and then run to failure at a higher stress level S_2, the remaining life at the stress S_2 is usually greater than the 40% cycle ratio predicted from Miner's Law. If the specimen is stressed first at the level S_2 and then run to failure at the lower stress S_1 the remaining life is found to be less than predicted. This result is general for steels and aluminium alloys.

Recently, some work has been published (5-7) on the propagation of macrocracks (i.e., 0.1 to 1.0 inches long) in wide sheet specimens (4 to 12 inches wide). Here the crack is many grains long, propagates in a direction normal to the direction

of maximum tensile stress and on a plane or planes 45° to the sheet surface. Crack propagation rates with increasing or decreasing stress have been measured and a result opposite to the above is indicated. It has been reported that when the test stress is decreased a period of very slow or even zero crack growth occurs. This has been taken to indicate a Miner's sum greater than unity. Conversely, with increasing stress the crack is reported to immediately take up a growth rate characteristic of the higher stress [5,6] or perhaps slightly greater[7]. The indicated cumulative damage sum in this case would be slightly less than or equal to unity. It is believed that such data as have just been discussed are fundamentally different than the bulk of the available cumulative damage data and the fact that the Miner's sums are indicated to be in the opposite sense, albeit disturbing, merely proves either that the methods of gathering data are not compatable or that the metallurgical and physical situations are different.

There are several other factors that directly invalidate the basic assumption of Miner's Law (that damage at a stress is proportional to the number of cycles at that stress) and which will clearly require consideration before a complete and unequivocal explanation of the failures of Miner's Law can be advanced. These are:

a. The number of cycles required to initiate a crack of a given length is a function of stress.

b. The rate at which that crack would grow if the stress were not changed is not a linear function of the number of cycles applied. Two types of crack growth have been found to occur; these are not the same function of the stress.

c. The influence of the crack density gradient, residual stresses, and stress relaxation on the rate of continued crack propagation when the stress is increased or decreased.

d. The fact that more total crack penetration can be tolerated prior to fracture as the stress is lowered.

It is proposed that the first three of the above mentioned difficulties in the understanding of cumulative damage fatigue may be investigated using a single research technique; the determination of crack density (damage) versus depth profiles in partially fatigued specimens. To do this one determines a fairly complete S-N curve, then cycles specimens for varying proportions of the total expected fatigue life (at two chosen stresses) and then counts the number of cracks (and measures their length) on the constant stress section of the surface. Following this the specimens are electropolished to remove a few microns and another crack count made, and so on to obtain a plot of crack density (or damage) versus depth.

Items a and b, above, can be quantitatively determined from the crack density vs depth data, and item d may be ascertained from the final fracture surfaces of the fatigue specimens or

from another series of specimens used for this purpose alone. Item c may be determined by applying various cycle ratios at S_1 and S_2, changing stress and applying a small cycle ratio (5 or 10 percent) and then doing a crack density vs depth survey as was described above. From these data it can be determined if the initial rate of crack propagation at the second stress is more or less than predicted.

It is believed that the effect of the damage profile set up at a given stress will be to alter the initial crack propagation rate at a second stress in such a way as to cause the disturbing variations in the cumulative cycle ratio at failure. Consider cumulative damage tests at the two stresses previously mentioned. For the case where a specimen is stressed first to 60% cycle ratio at a stress S_1, and then run to failure at a higher stress S_2, it is believed that there may be a period of incubation (where no crack growth occurs) or a period of increasing growth rate (during which the growth rate is increasing to that characteristic of the higher stress) while the crack density profile characteristic of the higher stress is created. Thus it might be expected that the sum of the cycle ratios would tend to be greater than one. If the specimen is stressed first at the higher level S_2, and then run to failure at the lower stress S_1, it is anticipated that there will be a period of faster than normal or a decelerating crack growth rate (during which the growth rate is decreasing to that characteristic of the lower stress) while the crack density vs depth profile adjusts itself to that characteristic of the lower

stress, the concommitant result being a cycle ratio sum tending to be less than unity.

EXPERIMENTAL

From the above discussion it is clear that the ideal material for this research would be a well work hardened one so that the elastic range is maximized, tested at stresses not too high so as to minimize or eliminate changes in matrix hardness and structure due to varying the applied stress. Ideally then, the fatigue study should be carried out well below the yield stress. For this research cold drawn commercial nickel was chosen. Its properties and chemistry are given in Table I below.

TABLE I

Hardness = 225 DPH

Yield Stress at 0.05% offset = 76,000 psi

Ultimate Tensile Strength = 103,000 psi

% Elongation = 15% % Reduction of Area = 36%

Average Grain diameter 75-90 microns.

C	0.054 wt%	Cu	0.03	Ti	0.02
N	.001	Fe	.13	Co	0.49
O	.0008	Mn	.19	Mo	< .005
H	.0005	Mg	.10	Al	.02
Si	.05	Cr	.02	B	.003

The fatigue testing was accomplished on deeply waisted cylindrical specimens having a 0.200 inch long constant stress section 0.250 inch in diameter. All specimens were electropolished prior to testing. The tests were performed in a standard six ton Losenhausen hydraulic fatigue machine at 1500 cycles per minute in alternating tension-compression about zero mean stress.

An S-N curve was roughly established and from this two stresses were chosen for emphasis and a fairly rigorous statistical determination of the mean life and scatter was then accomplished at the two chosen stresses. Note that even the higher stress is considerably below the static yield stress. The results at the two chosen stresses are:-

High Stress: 60,000 psi $\bar{N} = 7.324 \times 10^4$ cycles $\text{Log}\,\sigma = 0.0404$

Low Stress: 47,000 psi $\bar{N} = 3.143 \times 10^5$ cycles $\text{Log}\,\sigma = 0.0723$

where $\bar{N}$ is the mean fatighe life at the indicated stress and σ is the standard deviation.

The general hardness after fatigue was unchanged although microhardness tests did indicate some work softening on the fracture surface. In general, though, it is apparent that matrix stability has been achieved and fatigue results at the two chosen stresses should be comparable.

As has been mentioned the primary aim of this investigation was to measure and rationalize the damage occurring at several stages in cumulative damage fatigue. Crack density versus depth

measurements were performed following tests to 40, 50, 60 and 70% of the mean fatigue life in single stress fatigue. In addition cumulative damage determinations have been done for 40 + 20%, 40 + 30%, 50 + 10% and 50 + 20% tests with the higher stress applied first in half the cases, and vice versa. At this time only the 50 + 10% tests are sufficiently far along to present in this paper. Finally to avoid the eccentricity and spurious effects around inclusions usually associated with continued electropolishing each specimen was machined in a lathe and then polished with several grades of emery paper between electropolishing treatments.

The specimens were mounted in a fixture which allowed their free rotation and the specimen and fixture placed on the stage of a Bausch and Lomb Research Metallograph. Counting was done at 100X at which magnification a length of 0.050 inch could be scanned. The cracks within this length increment and completely around the circumference were counted and measured. Then the specimen was indexed 0.050 inches and another circumferential count made, and so on until all the cracks on the 0.200 inch constant stress section of the specimen had been counted. At greater depths in the specimen and at any time when there was any doubt about the existence of a crack, a check was made at higher magnification, either 250X or 400X, before going on.

Results and Discussion

The crack density vs depth data for the constant stress amplitude tests are shown (to a depth of 250 microns) in Figures 1 and 2. The first thing to be noted is the large number of cracks present on the specimen surface. Since the actual number of cracks on about one sixth square inch of surface were counted this means the crack density on the as fatigued specimen surface varied between about 6000 to 40,000/inch^2 for the lower stress to well over 1 x 10^5 cracks /inch^2 for the higher stress. It is important to note, in addition, that since the surface counts were so high only cracks longer than about one half the grain diameter were counted (this was not true below the surface; there all cracks were counted.) Therefore the surface crack density may be as much as an order of magnitude higher than reported here. Examination of taper sections has confirmed a very high density of surface cracks; no slip is observed.

It may be seen, as well, that with increasing cycle ratio these curves tend to shift upward and to the right, indicating that cracks, in general, tend to propagate once formed. If it is remembered that the average grain diameter is 75 - 90 microns it becomes obvious that by far the largest change in the number of cracks present occurs within the first grain layer inward. In fact, it would seem that even up to 40% of the fatigue life at the higher stress cracks would appear to have not yet penetrated a single grain boundary. The curves through the data are not smoothly drawn at the lower densities, since in many

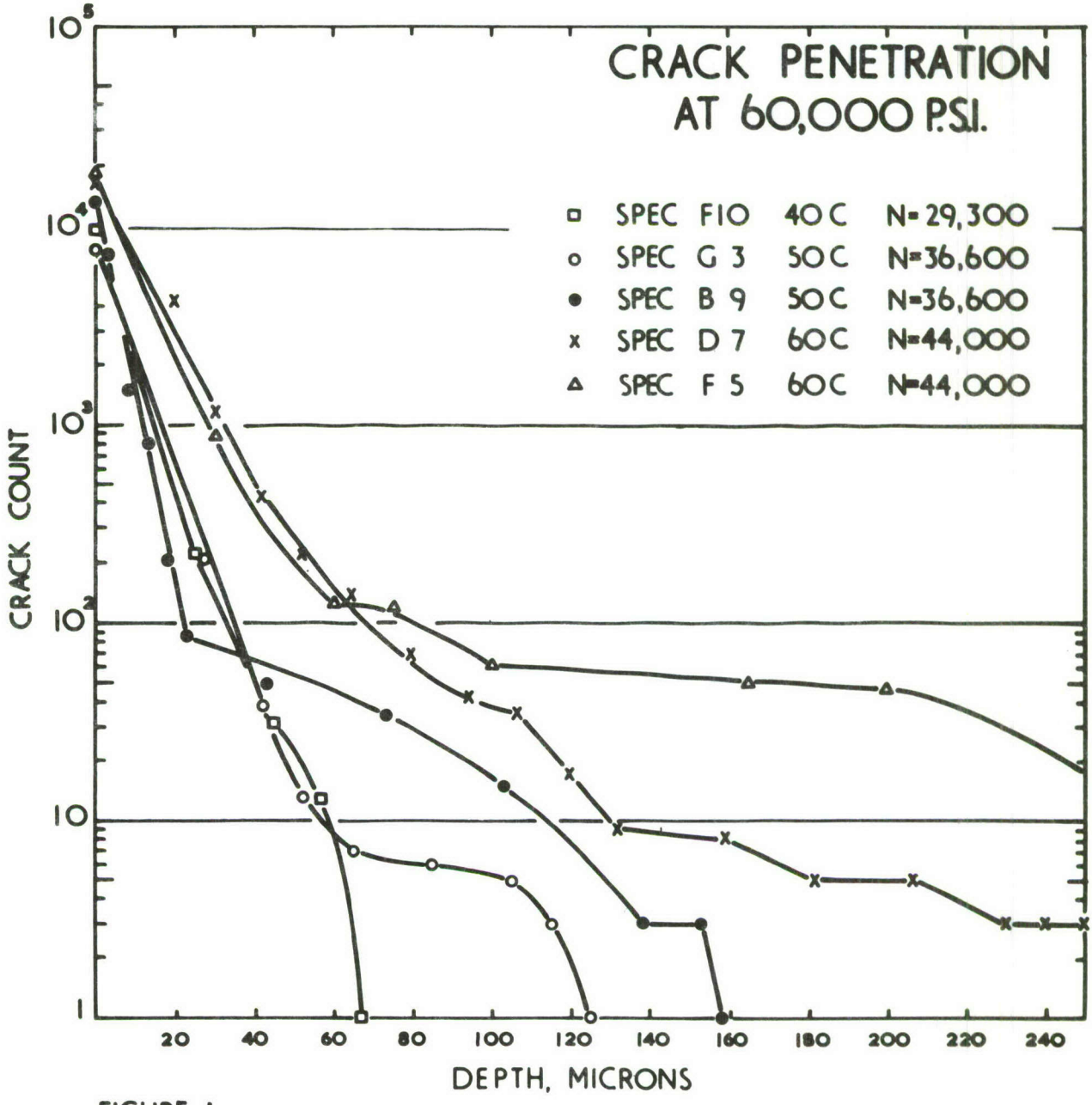

CRACK PENETRATION
AT 60,000 P.S.I.
SPEC F10 40C N=29,300
SPEC G 3 50C N=36,600
SPEC B 9 50C N=36,600
SPEC D 7 60C N=44,000
SPEC F 5 60C N=44,000
CRACK COUNT
10^5
10^4
10^3
10^2
10
1
20
40
60
80
100
120
140
160
180
200
220
240
DEPTH, MICRONS

FIGURE 1

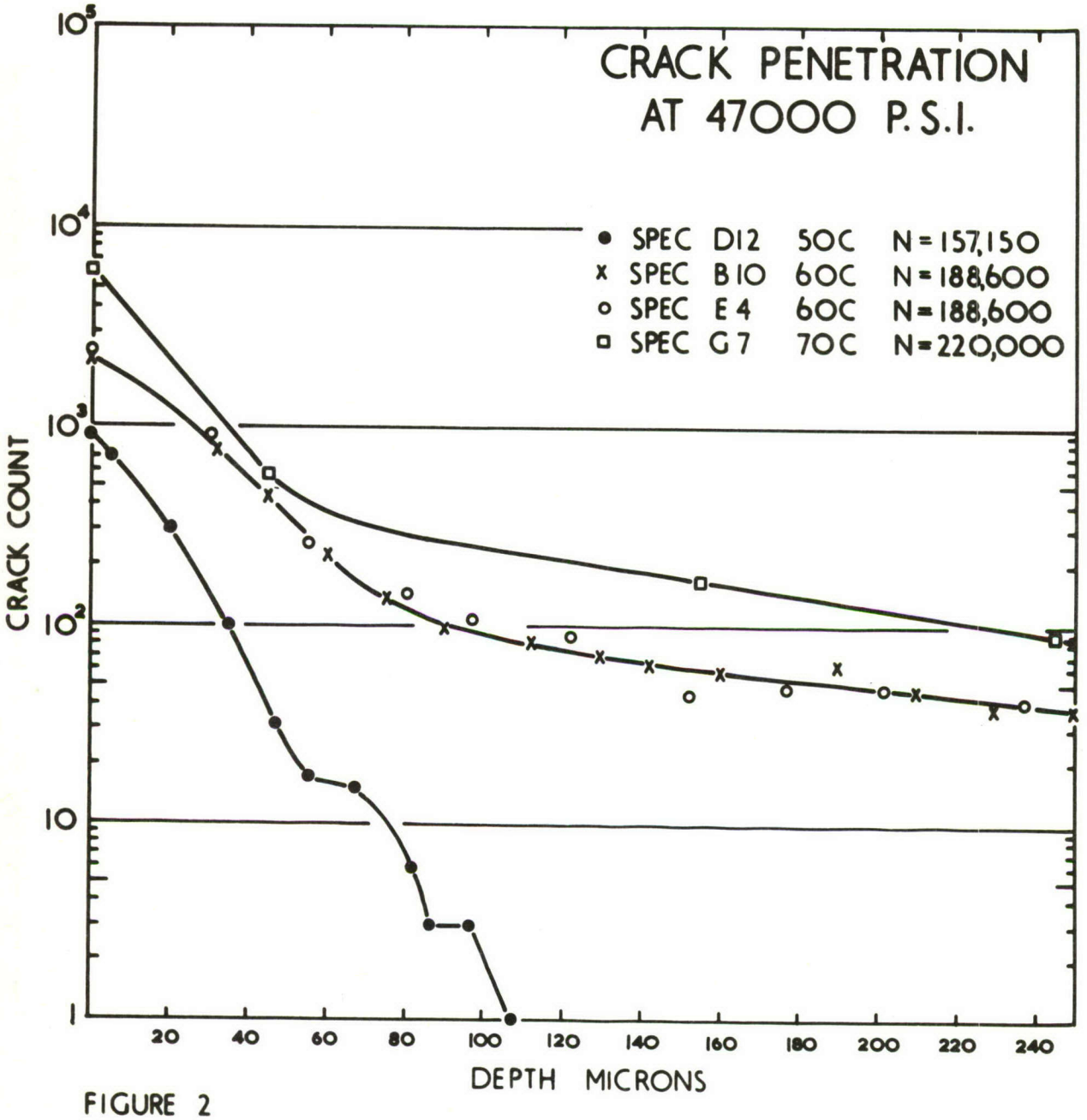

CRACK PENETRATION
AT 47000 P.S.I.
● SPEC D12 50C N=157,150
x SPEC B10 60C N=188,600
o SPEC E4 60C N=188,600
□ SPEC G7 70C N=220,000
CRACK COUNT
10^5
10^4
10^3
10^2
10
1
20
40
60
80
100
120
140
160
180
200
220
240
DEPTH MICRONS

FIGURE 2

cases the change of crack density with depth was discontinuous.

The microscopic observations of these specimens reveal several interesting features. Figures 3, 4 and 5 are of specimen D7 tested to 60% of the mean life at the higher stress. Figure 3 shows a typical view of the surface of an as-tested specimen. The size of the individual crack groups gives an idea of the grain size. It is relatively easy to imagine how these groups link to form a quite long surface crack. Figure 4 is from the same specimen but at a higher magnification. Here one can see the crack size distribution within a group, a few cracks nucleating at inclusions and also a few fatigue extrusions, and from this picture it would appear that all the surface markings are cracks; slip lines are not seen.

Figure 5 from the same specimen and at the same magnification as Figure 3 shows the dramatic decrease in crack density at a depth of only 20 microns. Note the dense black open appearance of these cracks nucleated at the surface; this will become important later. Figure 6 is of a cumulative damage specimen (G11, discussed later) but is shown here as being typical of the surface of a specimen tested to 60% or 70% of the mean life at the lower stress. The striking effect of stress on the surface appearance of a fatigue specimen is clearly seen by comparing this Figure with Figure 3. It is significant that the surfaces of the specimens tested for varying cycle ratios at a given stress all appear under the microscope to have roughly the same crack

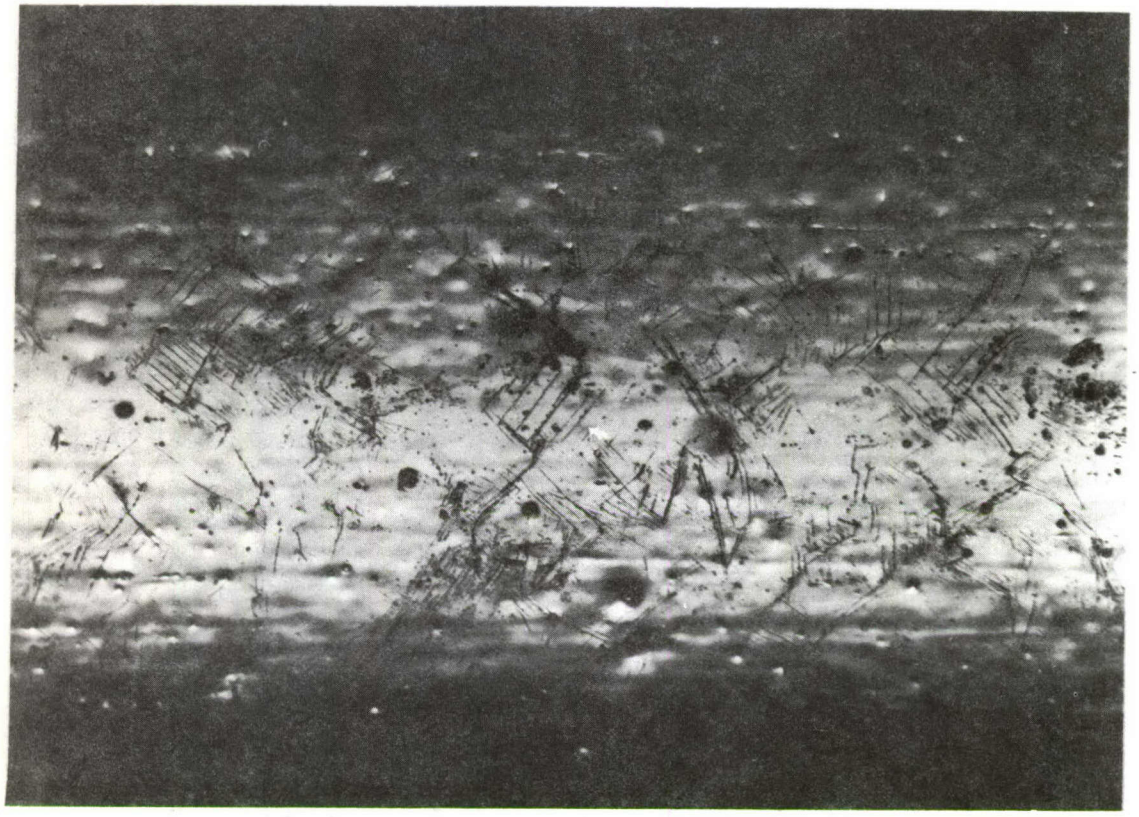


Figure 3. As tested view of specimen D7 tested to 60% cycle ratio at 60,000 psi, 125X

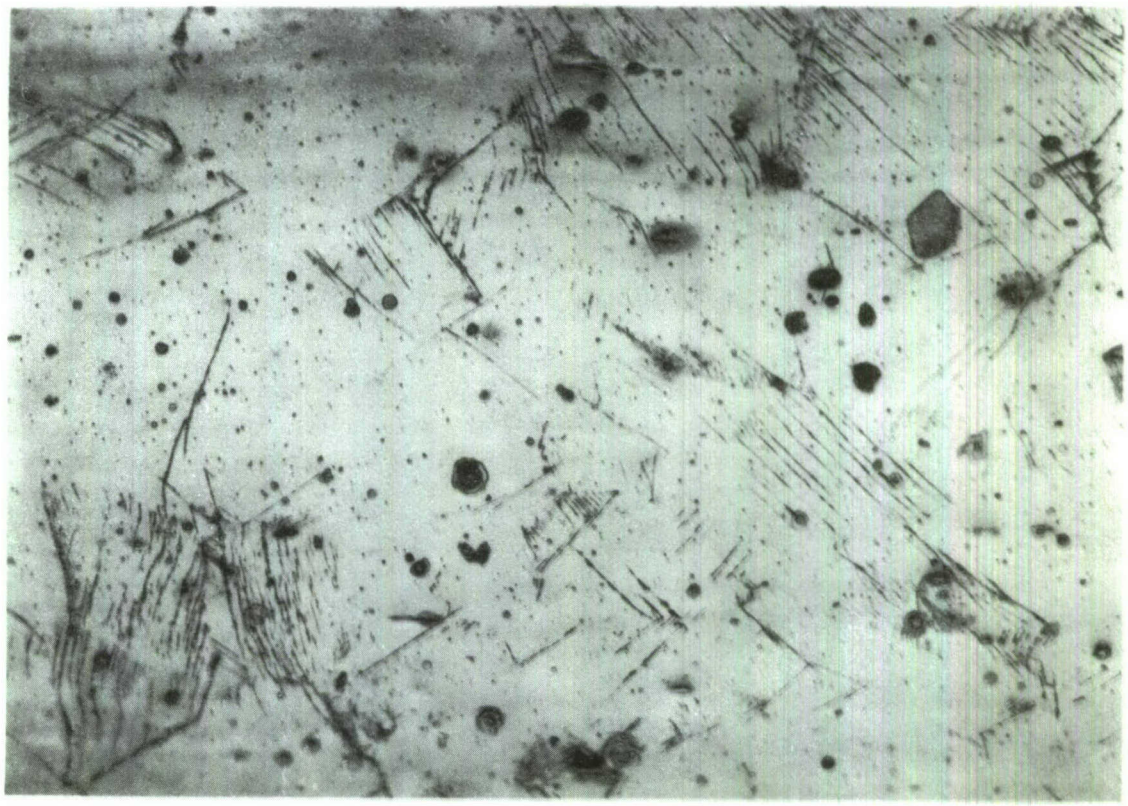


Figure 4. Surface crack groups, specimen D7 tested to 60% cycle ratio at 60,000 psi, 250X

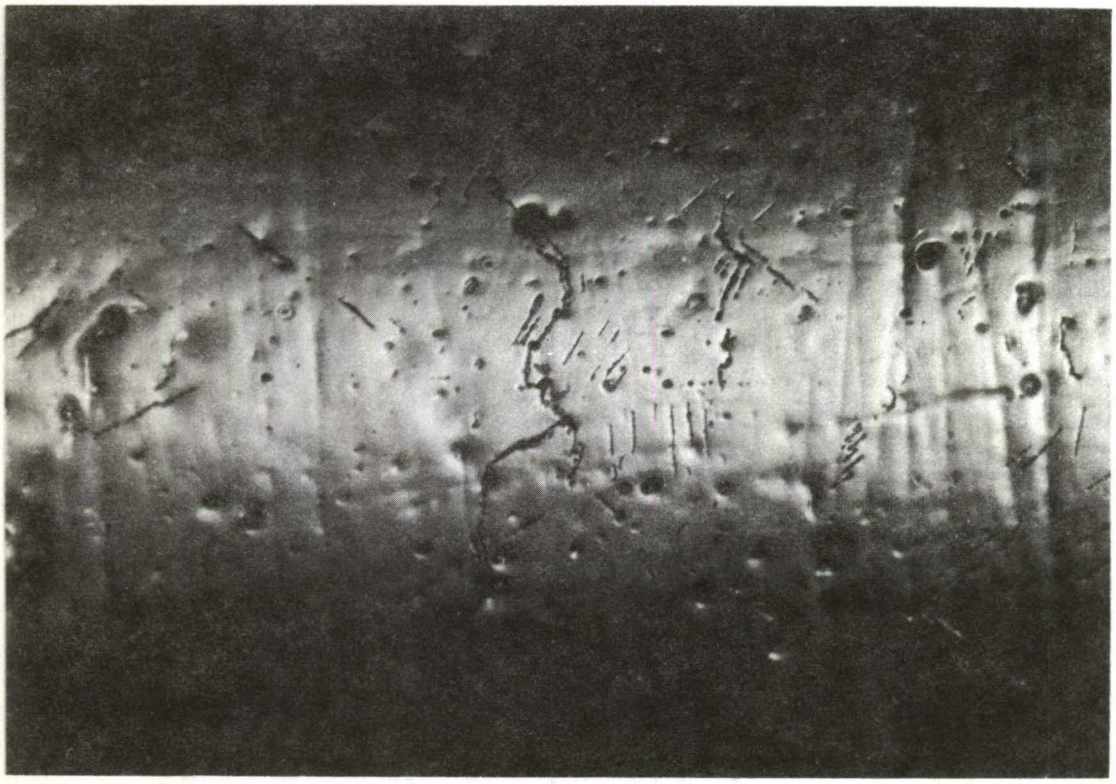


Figure 5. Subsurface view of specimen D7, tested to 60% cycle ratio at 60,000 psi, depth 20 microns, 125X

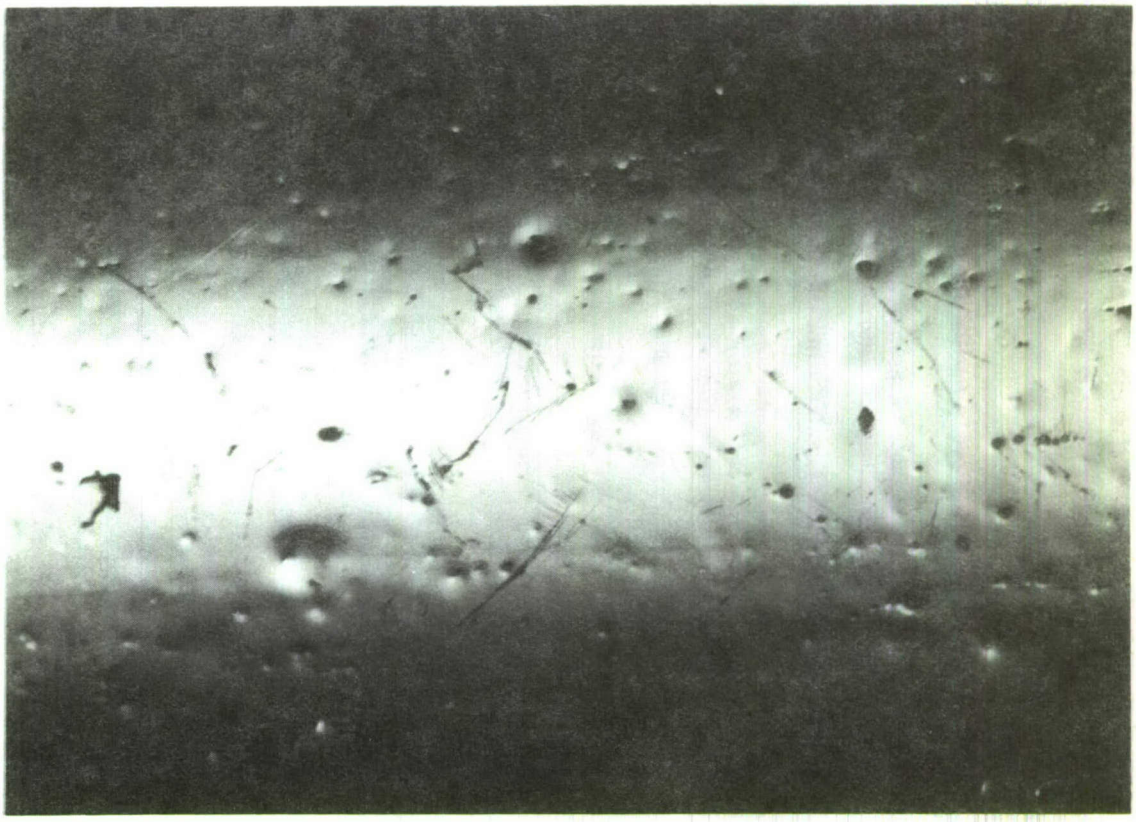


Figure 6. As tested view of specimen G11, tested to 50% cycle ratio at 47,000 psi plus 10% cycle ratio at 60,000 psi, 125X

density. Yet when the cracks are actually counted there seems to be a relationship between surface crack density and cycle ratio. The apparent explanation is that cracks which are evident but too tiny to count at one cycle ratio may continue to grow so that at some stage they would become larger than one-half the grain diameter and would then be considered.

The microscopic examination of these specimens leads to the general conclusions that;

a. all cracks are small (< 100 microns deep) at 40% or less of the expected mean life (hereinafter abbreviated as 40C)

b. Some of the cracks are larger at 50C but there is still almost no crack linking.

c. At 60C there is considerable crack linking at both stresses.

d. At 70C there is a relative absence of tiny cracks.

To return specifically to Figure 2 for a moment there are several features worthy of attention. First, the lack of scatter between the data of specimens B10 and E4 is remarkable, especially in view of the fact that B10 had a single huge fatigue crack with a total surface length of about one-third the circumference. This is the most startling example of a fact which makes itself apparent many times over in these data: that the presence of one or more large and deep fatigue cracks in a specimen does not seem to affect the growth of other cracks in the vicinity. The belief that only one crack penetrated to any significant depth is one of the

"old wives tales" of fatigue. That the "tale" is false is shown by the long plateaus in Figures 1, 2, 7 and 8, or by the fact that in the area immediately adjacent ($\pm$ 200 microns) to the very large crack in specimen B10 twelve cracks persisted to a depth of 1000 microns and eight of these still existed at a depth of 1400 microns.

Secondly, it should be noted that in several cases the data indicate the crack count to increase slightly with increasing depth. This is not due to scatter in the data but rather to the fact that some of the cracks split into several pieces with increasing depth, indicating that propagation is not uniform at these depths. It may be noted here that the cracks would all appear to be considerably longer than they are deep, i.e., they are roughly saucer shaped.

Thirdly, it may be observed that at some representative depth, as 200 microns, the specimens tested at the lower stress have a higher crack count than those tested at the higher stress. This does not necessarily mean that the lower stress is the more damaging, but rather that specimens tested at the lower stress have a larger number of small cracks whereas at the higher stress the specimens exhibit fewer larger cracks. This fact led to an actual measurement of the apparent crack length at various depths for several specimens, the results of which are reported later.

Figures 7 and 8 show cumulative damage comparisons to a depth of 1250 microns. Each graph shows lines of smoothed data

previously discussed (in Figures 1 and 2). In each figure the curves B show the scatter in crack density vs depth expected at 60C. Finally, the symbols show the actual data points for cumulatively damaged specimens. Figure 7 shows a specimen cumulatively damaged by a 50% cycle ratio at the high stress (now to be abbreviated as 50H) followed by a 10% cycle ratio at the low stress (now to be abbreviated as 10L). It is compared to two specimens tested to 60H. In addition, the deepest penetration at 50H is shown to indicate the exact influence of the 10L. This graph shows that, as expected, the surface and immediate subsurface density of cracks in the cumulatively damaged specimen compare very well with the data from 60H. It is apparent that the main effect of the final 31,400 cycles at the low stress has been to considerably extend the dominant crack while not greatly affecting the rest of the crack density-depth relation.

Figure 8 shows the data for the specimen cumulatively fatigued to 50L + 10H. Again for comparison the data for 50L and 60L are shown. For the condition 50L + 10H two effects are noted. First, the surface crack density has started to increase (Figure 6) indicating a marked difference caused by the order of application of the stresses. Second, the 7300 cycles applied at the higher stress have resulted in a very significant penetration of the dominant cracks, although the last 10H has caused less penetration than would a final 10L.

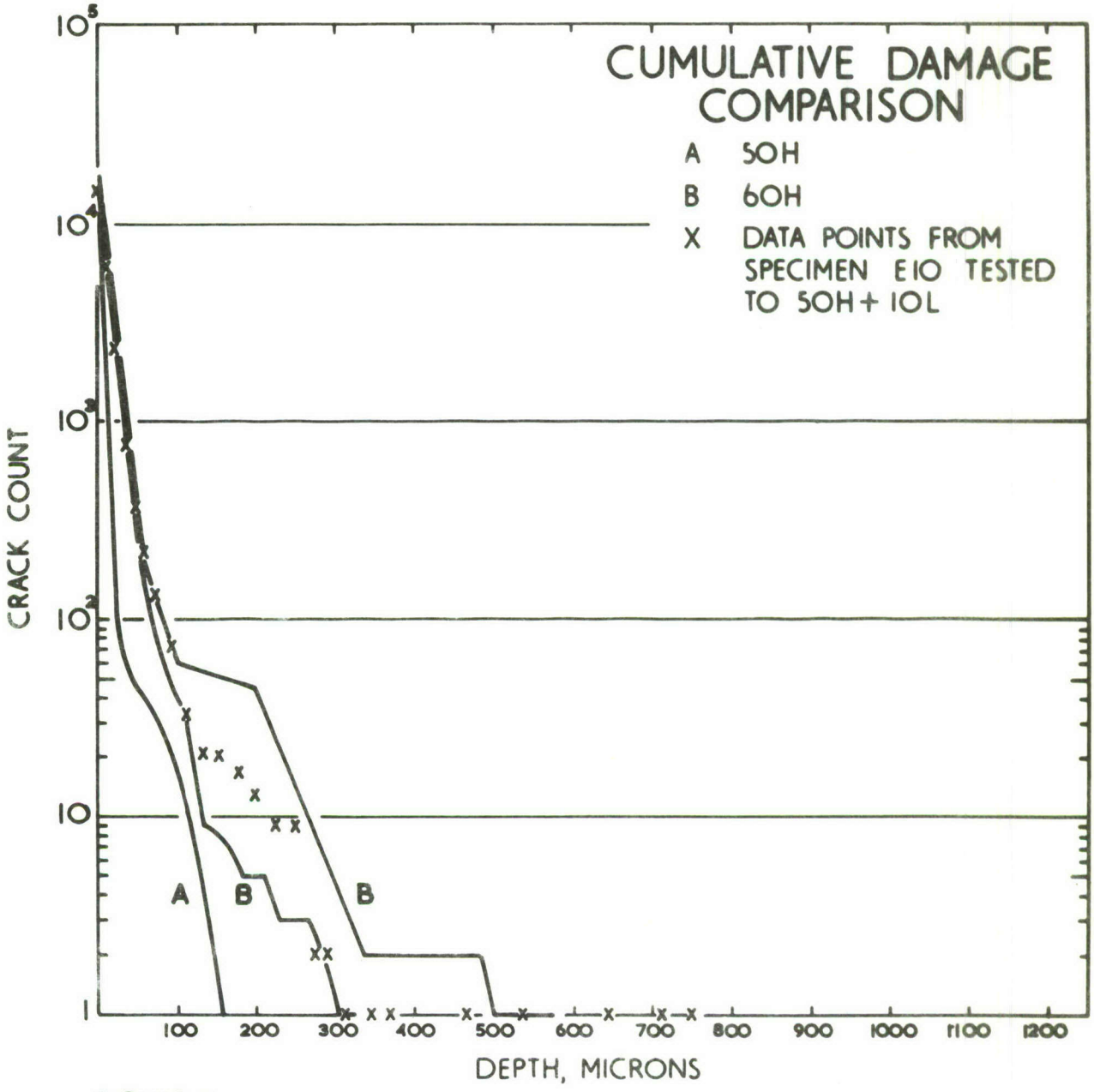

CUMULATIVE DAMAGE COMPARISON
A 50H
B 60H
X DATA POINTS FROM SPECIMEN E10 TESTED TO 50H + 10L
CRACK COUNT
DEPTH, MICRONS
A
B
B

FIGURE 7

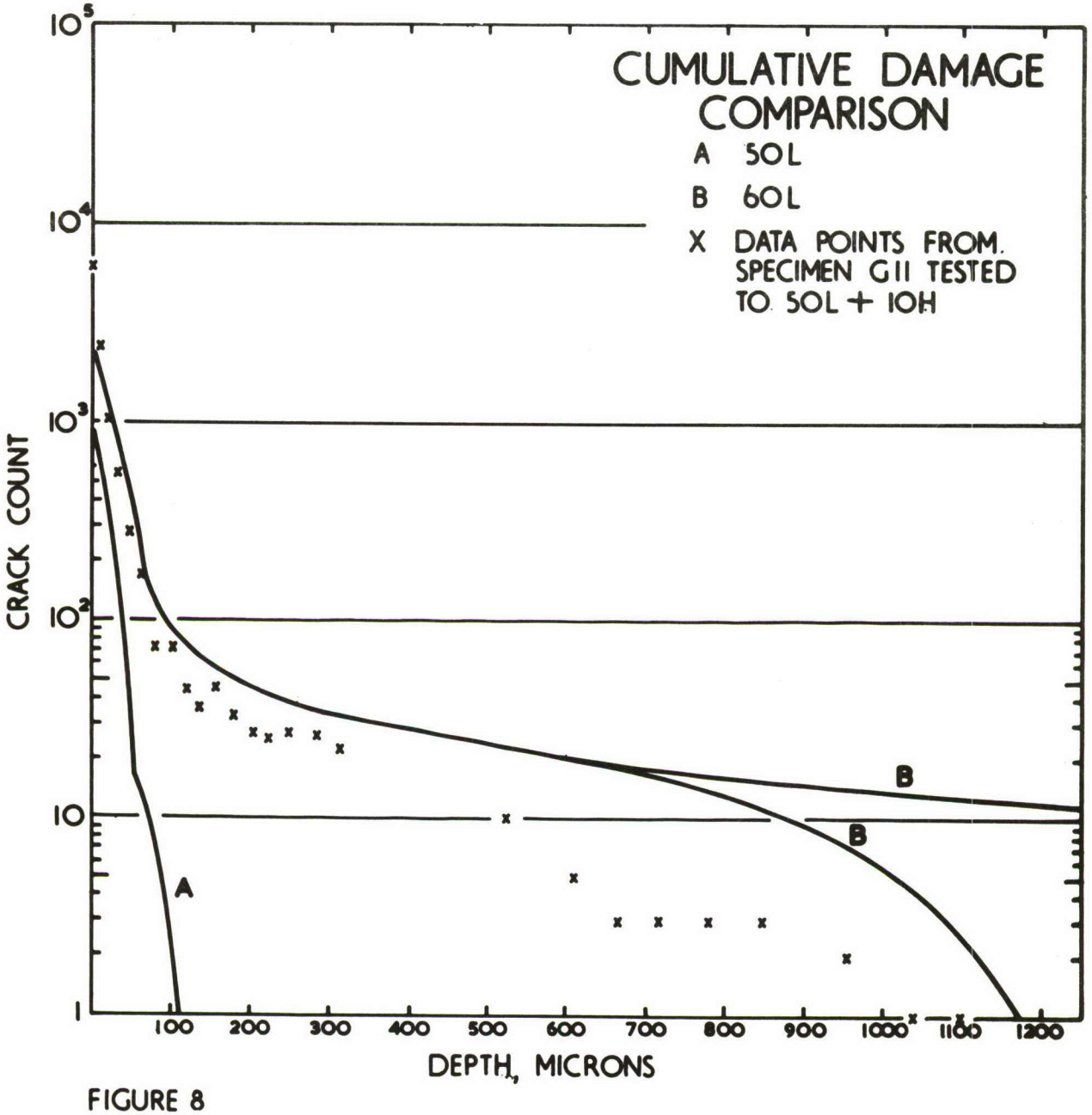

CUMULATIVE DAMAGE COMPARISON
A 50L
B 60L
X DATA POINTS FROM. SPECIMEN G11 TESTED TO 50L + 10H
CRACK COUNT
DEPTH, MICRONS
A
B
B
10⁵
10⁴
10³
10²
10
1
100
200
300
400
500
600
700
800
900
1000
1100
1200

FIGURE 8

It is apparent, though not yet unequivocal, that these data will be able to explain the failures in Miner's Law. For the result of the high-low sequence would appear to lead to a Miner's sum of less than unity (Figure 7), while the result of the low-high sequence would indicate a damage sum of greater than unity (Figure 8). Such is exactly the behavior of this material when cumulative damage tests are carried to failure. It is also apparent that the factor leading to this conclusion is that in the high-low sequence the dominant crack seems to propagate further at this stage of fatigue by the application of 10L than by 10H. The reverse is also true; for the low-high sequence the dominant crack would seem to be propagated less by 10H than by 10L.

Before leaving these data it is instructive to consider the actual rates of crack propagation. Table II shows these data for Figures 7 and 8.

TABLE II

	Increment	Penetration	Cycles	Growth Rate
1.	0-50H	160 microns	36,600	0.0044 microns/cycle
2.	50H-60H	300	7300	0.041
3.	50H-50H+10L	600	31,400	0.019
4.	0 - 50L	100	157,000	0.00064
5.	50L - 60L	1400	31,400	0.045
6.	50L-50L+10H	1000	7300	0.14

These data indicate that in the first 50% of fatigue life the average growth rate of the dominant crack is seven times as fast at the high stress as at the low (compare lines 1 & 4). But between 50 and 60% of the lives the average growth rates are approximately equal (lines 2 & 5). However, these data indicate that the effect of adding 10L to 50H (lines 2 & 3) is to cut the rate of propagation in half, but to allow this to go on for four times as long, the result being about twice as much propagation as would have occurred if the stress had not been lowered! Conversely, the effect of adding 10H to 50L (lines 5 & 6) is to increase the propagation rate by a factor of three, but to only allow this high rate to continue for one-fourth the number of cycles; the resultant is then less propagation (about 70%) than would have occurred if the stress had not been raised! It is equally important to note that these data would also seem to support those data gathered by others who measured only the change in rate of growth on increasing or decreasing the stress [5,7]. This comparison is possible because the dominant crack lengths are essentially equal at 50H and 50L. Such a comparison shows that on raising the stress the rate of crack propagation is three times that which might be expected (compare lines 2 & 6), whereas on lowering the stress the rate is a bit less than half that which might be expected (compare lines 3 & 5). This, then, points up the fallacy in measuring a rate of crack propagation and assuming that the Miner's sums would be similarly affected. If Miner's Law

were valid it should also be permissable to compare the total depth of penetration for the cases of lines 3 and 5 or lines 2 and 6. It may be seen that these comparisons would also tend to indicate a reversal of the expected deviations from Miner's Law. The single fact emerging unequivocally from these comparisons is that one must proceed with extreme caution in attempting to analyse cumulative damage data.

It was mentioned previously that during this program it became evident that simply counting the cracks might be leading to the creation of a false impression of the true extent of the damage. Because of this the crack length was also measured and, in general, the total crack length at any depth has been found to correlate fairly well with the crack counting results at the same depth. The crack length data are incomplete since they have only been taken during the latter half of the program but, for example, the summation of crack lengths at 400 microns depth are in the ratio 6:30:14 for the lines 4, 5 and 6 in Table II.

All the data presented thus far have been as regards cracks which originated at the surface and propagated inward. Recently, however, Pelloux[8] reported finding evidence (in electron micrographs of fatigue fracture surfaces) of cracks apparently nucleated at inclusions within the specimen. He showed that in many cases these tended to speed up crack propagation markedly. What he was unable to show by electron microscopy was the large

number of internally nucleated "incipient" cracks and the mechanism of their nucleation.

These incipient cracks have been observed in this investigation in all specimens where crack growth extended to more than 300 microns depth. The seriousness of the problem may be seen in Figures 9 and 10 which are the full crack density vs depth plots for the cumulatively damaged specimens previously discussed. Here are shown the total crack count and the separate count of cracks nucleated at the specimen surface (called "original" cracks). The difference between these two is the number of incipient cracks apparent at any given depth. It may be seen that these first begin to appear at about 300 microns depth (although in a later specimen a few have been seen as early as 160 microns), reach a maximum at about 6-700 microns depth and continue for several hundred microns after the last "original" crack has vanished. An attempt has been made in all current specimens to differentiate between "original" and "incipient" cracks on the basis of their appearance.

Figures 11 and 12 show incipient cracks in specimen E4 at a depth of 620 microns. In Figure 11 the crack on the left is emerging from one side of an inclusion, while the crack at the right has propagated from both sides of the inclusion. The crack in the middle is an original crack which cracks often polish more rapidly than the matrix and appear to lie in the bottom of a (polishing) pit. The incipient cracks have a definitely

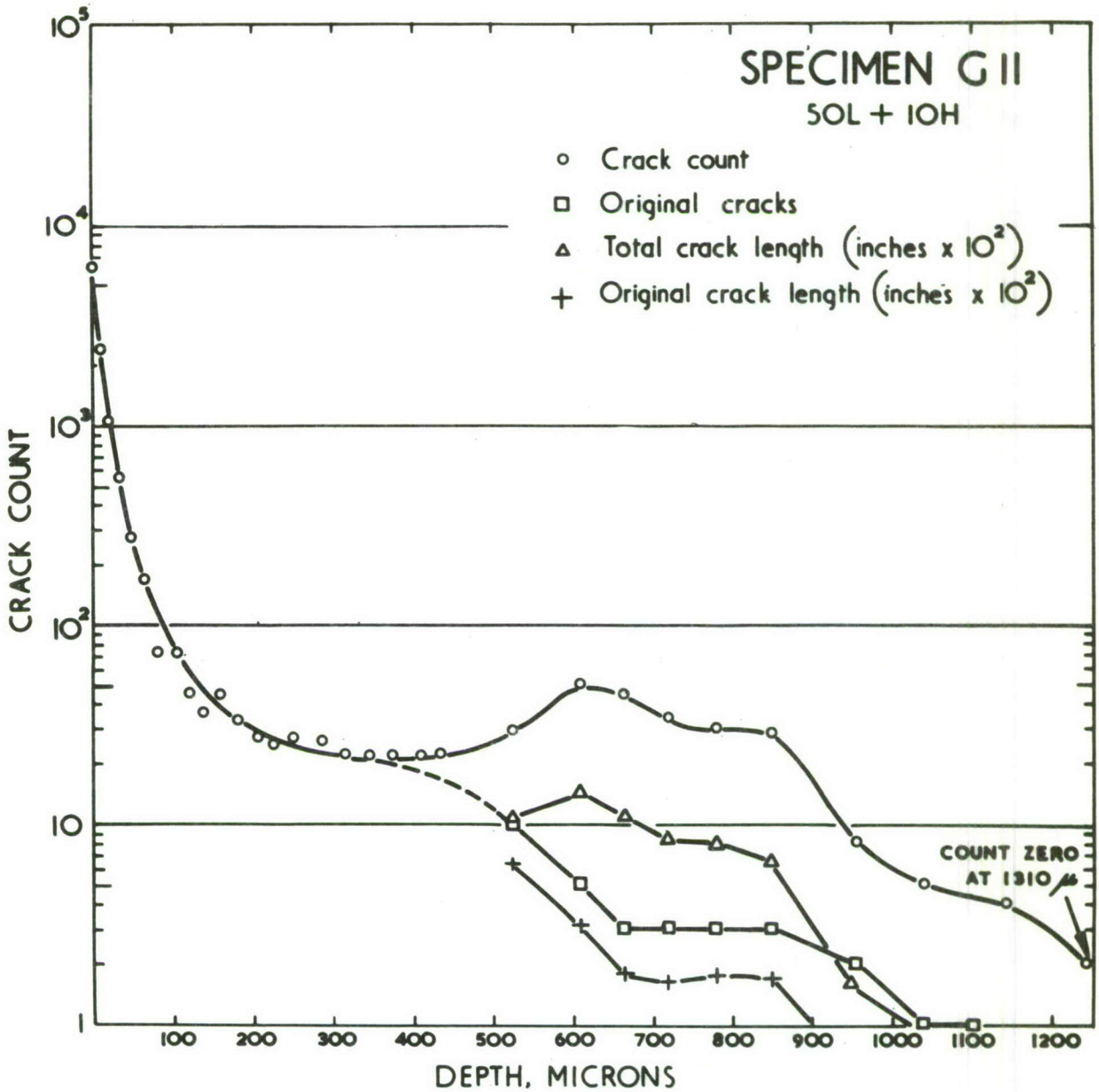

SPECIMEN G II
50L + 10H
Crack count
Original cracks
Total crack length (inches x 10²)
Original crack length (inches x 10²)
CRACK COUNT
10⁵
10⁴
10³
10²
10
1
COUNT ZERO AT 1310 μ
100
200
300
400
500
600
700
800
900
1000
1100
1200
DEPTH, MICRONS

FIGURE 9

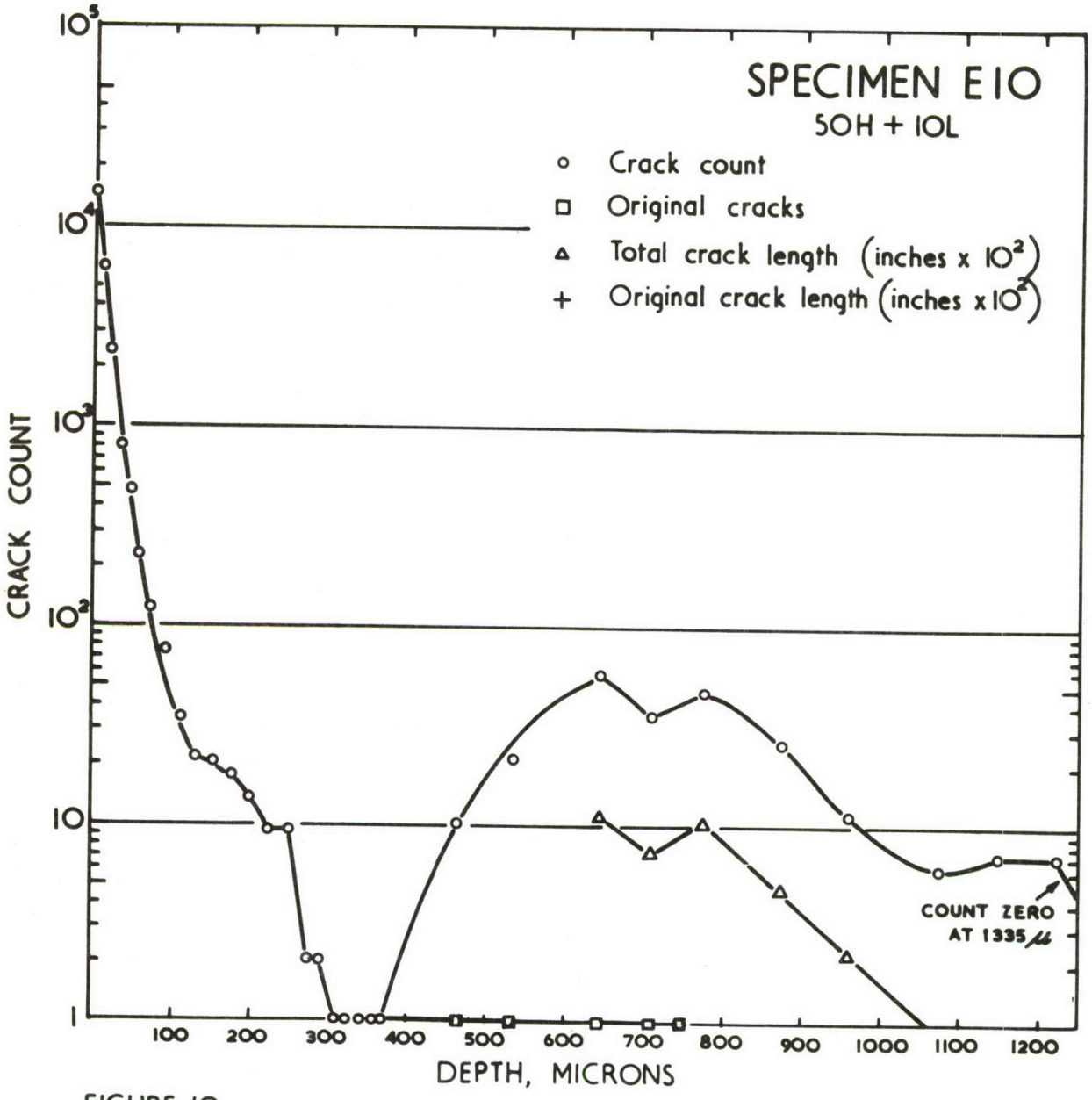

SPECIMEN E10
50H + 10L
Crack count
Original cracks
Total crack length (inches x 10²)
Original crack length (inches x 10³)
CRACK COUNT
DEPTH, MICRONS
COUNT ZERO AT 1335 μ
10⁵
10⁴
10³
10²
10
1
100
200
300
400
500
600
700
800
900
1000
1100
1200

FIGURE 10

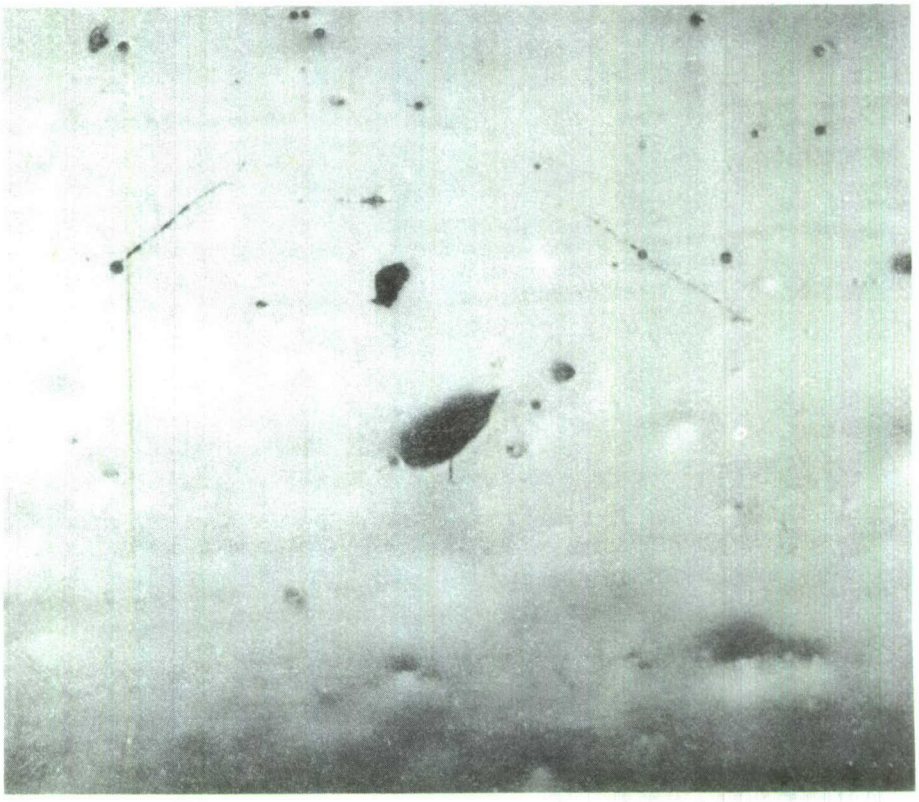


Figure 11. Two incipient cracks and an original crack in specimen E4, tested to 60% cycle ratio at 47,000 psi, depth 620 microns, 200 X

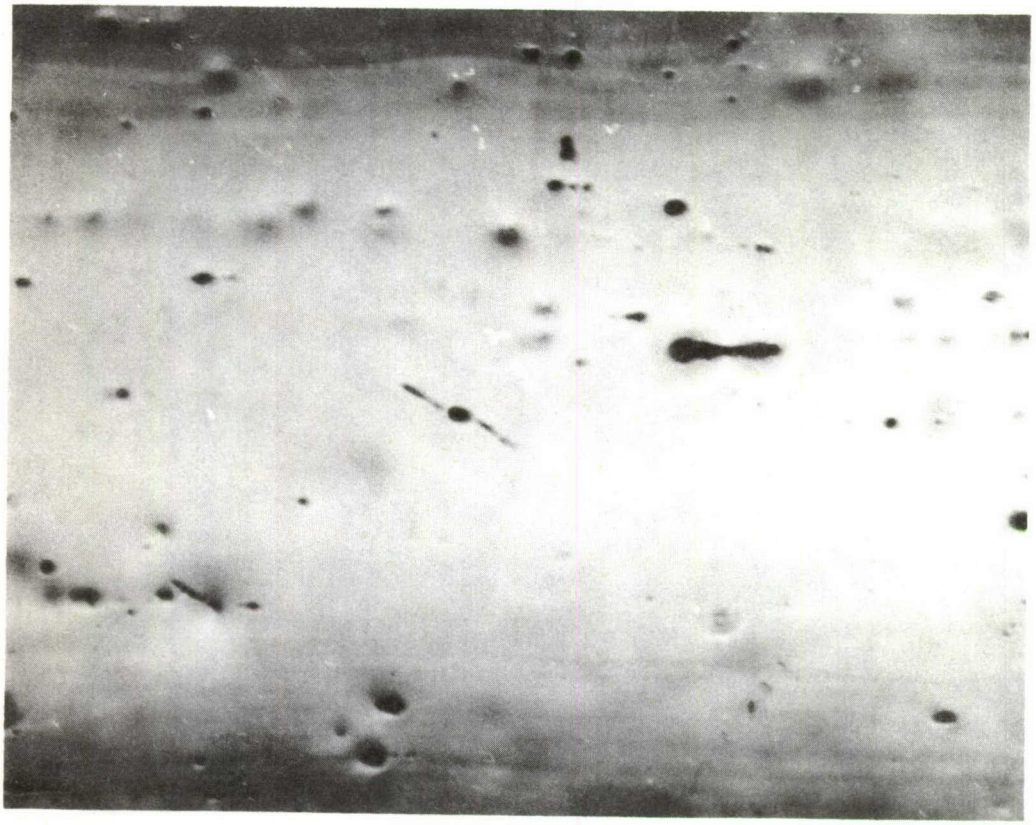


Figure 12. One incipient crack in specimen E4, tested to 60% cycle ratio at 47,000 psi, depth 620 microns, 200 X

different appearance and this has enabled them to be counted separately. It is notable that these cracks appear at about 45° to the specimen axis indicating that their formation is associated with a shear stress. It is believed that their formation occurs by the Zener mechanism[9]. During one half cycle a dislocation pileup occurs on one side of an inclusion with such rapidity that the normal process of cross-slip does not have time to eliminate the stress concentration before it becomes large enough to exceed the fracture strength of the material and a crack opens up perpendicular to the slip plane. On the reverse half cycle a crack could be created on the opposite side of the inclusion in a similar manner. The nature of their formation indicates that the two cracks should not lie on the same plane, but should appear to emanate from the edges of the inclusion as is shown pictorially in Figures 11 and 12 and schematically in Figure 13.

Several steps have been taken to verify that these cracks are not artifacts, that they are, in fact, generated by the fatigue stressing and to verify the proposed mechanism. An unfatigued specimen has been used at every stage along with the fatigued ones. It has been alternatively machined and polished at the same time as all the others and examined very closely by the same techniques. No markings of any sort have been observed on the unfatigued specimen to a depth of greater than 500 microns. This proves that the incipient cracks are not polishing artifacts and that they are definitely caused by cycles

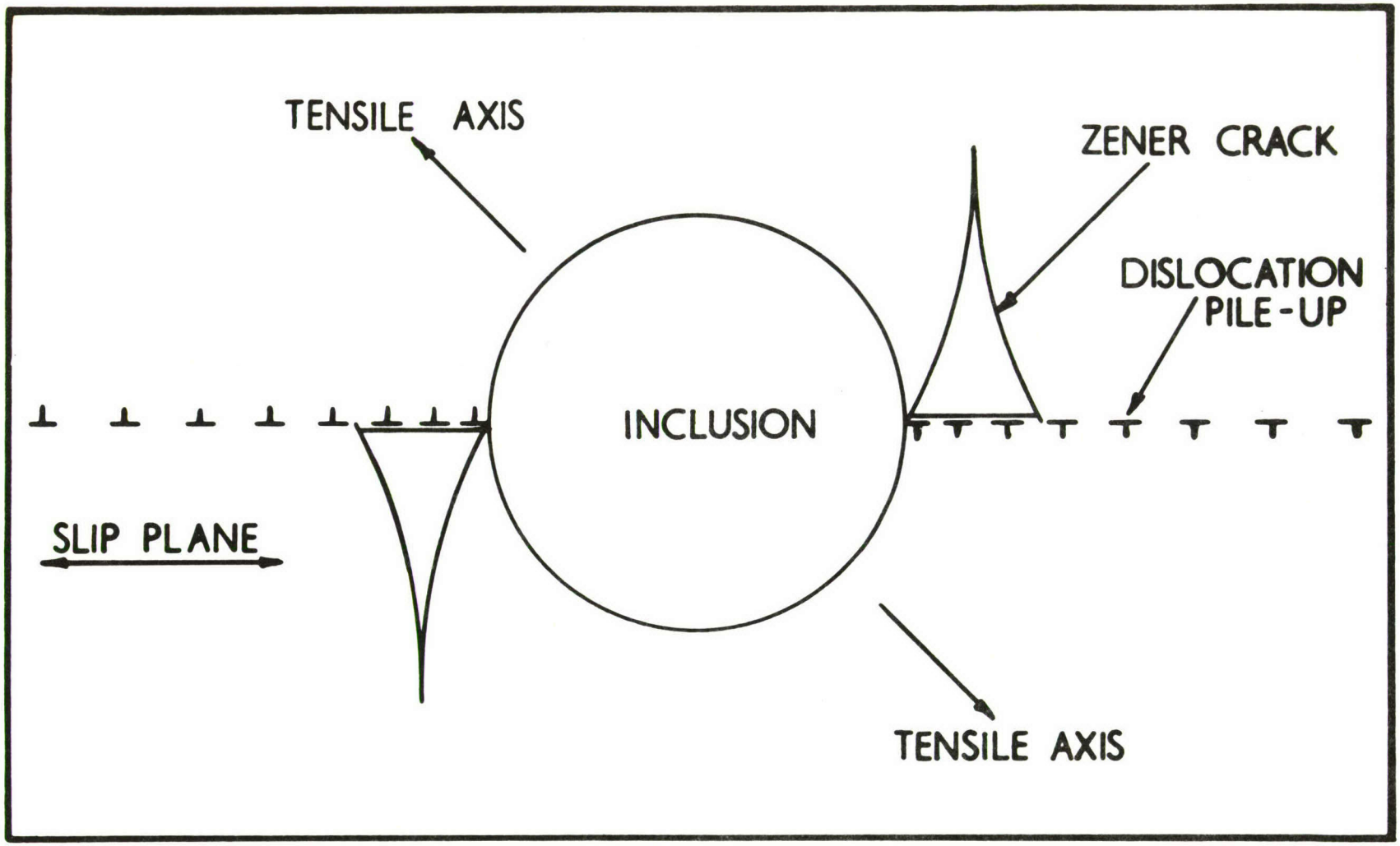


FIGURE 13 ZENER CRACK FORMATION - SCHEMATIC

of fatigue. The hypothesis that the incipient cracks are formed by a Zener mechanism is being investigated by back-reflection x-ray micro-beam analysis. This mechanism is also being studied in several other ways; the results of these investigations will be reported in a later paper.

SUMMARY AND CONCLUSIONS

This research has shown that several of the truisms of fatigue are false. It has been shown that smaller fatigue cracks continue to propagate even in the presence of a very large neighbor. It has also been verified that surface damage is primarily a function of stress and not cycle ratio, and that the vast proportion of fatigue cracks never get beyond the first grain boundary. As the cycle ratio increases an ever increasing number of cracks surpass the original barrier and move toward the center of the specimen. Crack linking is obviously favored at the higher stress where the higher resolved stress has enabled more grains to undergo cracking and the linking is done by short linkages. Thus, at some depth, there are fewer larger cracks formed at the high stress and a larger number of smaller cracks formed at the low stress. It is also clear that at least at the higher cycle ratios the damage at the low stress extends deeper than at the high stress; Miner's Law cannot, therefore, be expected to be valid. The reason for this is apparently that after the crack attains some length the stress concentration of the crack is sufficient such that small changes in stress are of little consequence.

A few cumulative damage experiments have indicated clearly the origin of the failures in Miner's Law and it is hoped that future results will allow the full rationalization of these failures. These data have shown that for a two step fatigue program the Miner sum is greater than unity, if the low stress is first applied, because the initial propagation of the dominant crack is apparently retarded at the high stress while the applied work goes to increase the surface crack density. Conversely, where the high stress is first applied, on lowering the stress apparently the dominant crack propagates immediately and to a greater depth than might be expected simply because the specimen is subjected to so many more stress reversals. These same data have shown the apparent fallacy in measuring crack propagation rates with stress change, and assuming that the Miner's sums would be similarly affected. Propagation rate results may also be derived from the present data but do not lead to a reversal in the expected deviations from Miner's Law.

Finally, a band of incipient fatigue cracks generated inside the metal have been herein reported for the first time. It has been shown that these cracks are caused by fatigue action at inclusions: the seriousness of this discovery cannot be minimized. In this very dramatic way the presence of inclusions has again been shown to be detrimental to fatigue properties. A mechanism for the origin of these cracks has been proposed and partially substantiated.

ACKNOWLEDGEMENTS

This research was accomplished during the author's sabbatical year in the Department of Metallurgy, University of Cambridge. The author wishes to thank Professor A.H. Cottrell, F.R.S., for having kindly made available to him the facilities of the Department of Metallurgy. The author wishes to express his appreciation to Mr. G.C. Smith (Cambridge), Dr. Geoffrey Chadwick (Cambridge) and Dr. Campbell Laird (Ford Motor Co.) and to his other colleagues who have provided him with continual inspiration and assistance. Mr. B. Goshawk and Mr. S.D. Charter performed the fatigue tests.

The author also wishes to acknowledge his debt to Mr. E.J. Hassell (ARZ) and Colonel R.E. Fontana, Director, ARL, and others who have made real the concept of a sabbatical year for Government employees.

REFERENCES

1. M.A. Miner, "Cumulative Damage in Fatigue", J. Appl. Mech., 12, p.A-159, 1945.

2. B.F. Langer, "Fatigue Failure from Stress Cycles of Varying Amplitude", J. Appl. Mech., 4, p.A160, 1937.

3. S.R. Valluri, "A Theory of Cumulative Damage in Fatigue", ARL 182, Aerospace Research Laboratories, Wright-Patterson Air Force Base, Ohio, 1961.

4. Many of the pertinent cumulative damage data are summarized in three recent books:
   a. P.G. Forrest, "Fatigue of Metals", p.114-119, Pergamon Press, 1962.
   b. J.A. Pope, "Metal Fatigue", Chap. 5, p.103-112, Chapman and Hall, Ltd., 1959.
   c. T.J. Dolan, "Basic Concepts of Fatigue Damage in Metals", Chapter 3 in "Metal Fatigue", McGraw-Hill,1959.

5. H.W. Liu, "Crack Propagation in Thin Metal Sheet under Repeated Loading", WADC TN 59-383, Aerospace Research Laboratories, Wright-Patterson Air Force Base, Ohio,1959.

6. H.F. Hardrath and A.J. McEvily, "Engineering Aspects of Fatigue Crack Propagation", Crack Propagation Symposium, Cranfield, p. 231, 1961.

7. R.H. Christensen, "Fatigue Cracking, Fatigue Damage and Their Detection", Chapter 17 in "Metal Fatigue", McGraw-Hill, 1959.

8. Regis M.N. Pelloux, "Influence of Constituent Particles on Fatigue Crack Propagation in Aluminium Alloys", D1-82-0297, Boeing Scientific Research Laboratories, September 1963.

9. C. Zener, "The Micro-Mechanism of Fracture", A Chapter in "Fracturing of Metals", ASM, p.3, 1948.

MARTIN McDERMOTT BRETTING
MAJOR USAF

WALTER ROALD NIESSEN
1/LT USAF

AIR FORCE ROCKET PROPULSION LABORATORY
EDWARDS AIR FORCE BASE, CALIFORNIA
(RPMC)

THE STUDY OF AN IGNITION TECHNIQUE FOR LARGE SOLID ROCKET MOTORS

(UNCLASSIFIED)

INTRODUCTION

The design of igniters for solid propellant rocket motors is largely an art. Despite recent developments which have increased our quantitative understanding of the propellant ignition process, it is not yet possible to confidently formulate equations predicting total motor ignition behavior. The interactions of the igniter energy release pattern, transient motor chamber pressurization and the bootstrap contribution of energy release by propellant ignition are, at the present, far too complex for quantitative analysis.

In the majority of igniter design programs, the inability to analytically design the igniter and predict ignition performance is not serious. Modern solid propellants are energetic and, in the general case, readily respond to the igniter's thermal flux. In these cases, the Art is sufficient to meet the needs of a rocket motor development program. On occasion, however, ignition problems arise for which the Art is inadequate. If discharge of the igniter initiates a cyclic chuffing behavior or fails to achieve stable ignition altogether, the Art's answer is, "Make it bigger." This often leads to high peak pressures which threaten the mechanical integrity of the motor case and propellant grain. This problem is usually solved by trial and error: a costly and time consuming process.

A more serious problem arises when the ignition technique, motor size or propellant grain design deviates from those previously encountered. If the deviation is great enough, one's confidence in

obtaining acceptable ignition behavior on the first try is lessened. The growing solid propellant motor ignition technology can provide only guidance for such an igniter development program. In order to gain sufficient confidence to use a new ignition technique, the igniter designer relies heavily on the determination of emperical design rules. These rules are derived through a number of subscale firings which include the variation of igniter and motor parameters which are presumed to control the ignition process.

One question always remains after such a characterization program: Did the testing program include and properly evaluate all the parameters which have profound effects on the ignition behavior? When one considers the ignition of a six million dollar large solid rocket motor, this question generates a great concern.

This paper describes the results of an in-house program accomplished at the Air Force Rocket Propulsion Laboratory concerned with an aft-end mounted ignition technique for solid propellant rocket motors. The stimulus for the work arose from questions on the mechanism and potential hazards of the ignition technique to be used in the Air Force's 156 inch and 260 inch diameter large solid rocket motor development programs. In these firings, a comparatively small solid rocket motor, or Pyrogen[1], is mounted in the expansion cone of the motor to be ignited (Figure 1). The hot gases generated by firing the pyrogen

[1]A small solid rocket motor used for ignition purposes was coined "pyrogen" by its originators, the Thiokol Chemical Corporation. This terminology has been accepted by most of the solid rocket community and will be used in this paper.

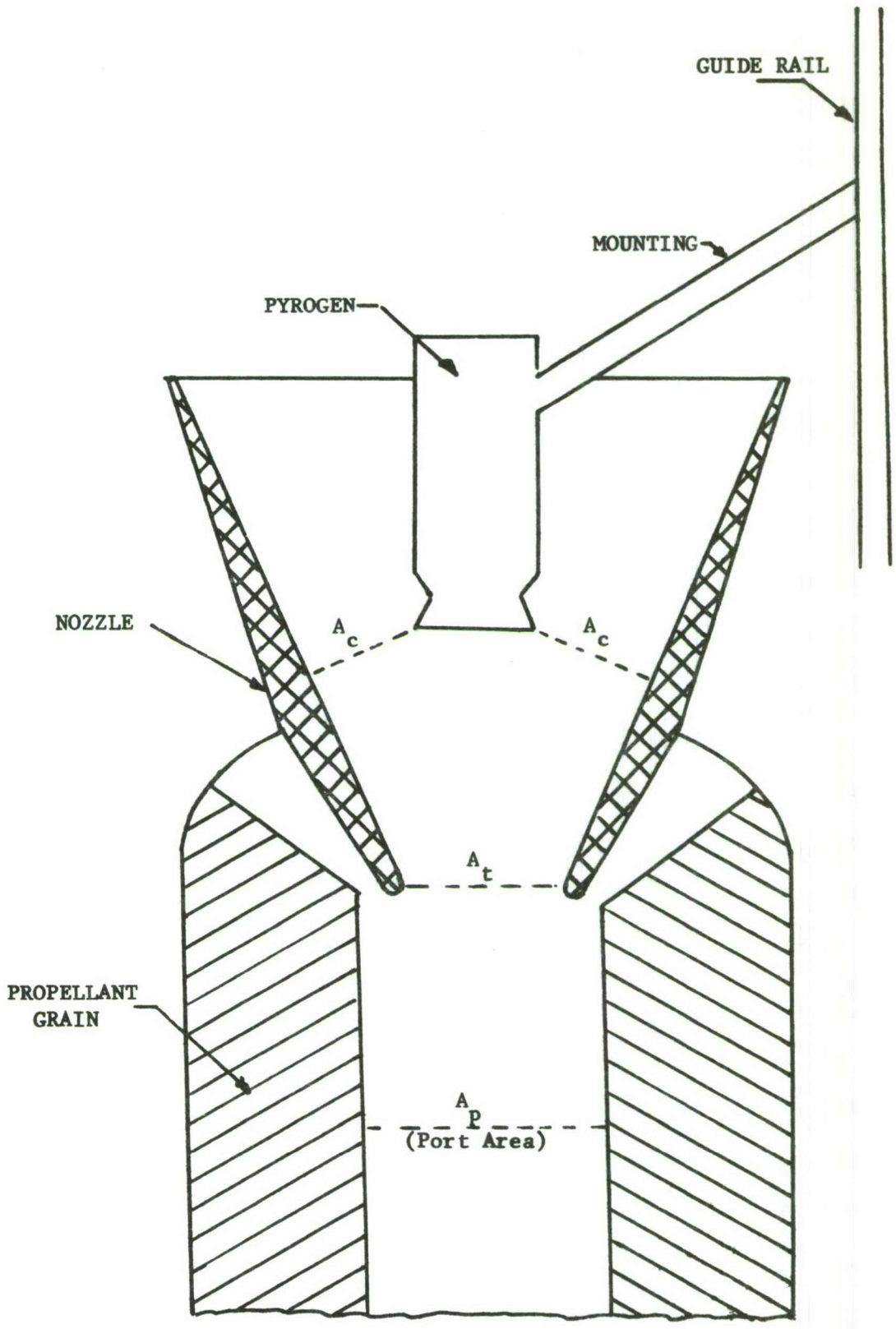


FIGURE 1 - SCHEMATIC OF AFT-END PYROGEN IGNITION SYSTEM

are discharged into the port cavity of the main motor, thus effecting its ignition. A pressure transducer, sensing a pre-set pressure level in the main motor, triggers a mechanism which releases the pyrogen motor assembly from its tie-downs. Under the action of its own thrust and the drag forces created by the main motor discharge, the igniter assembly moves away from the now ignited main motor along a short track. After reaching the end of the track, a cable restraint system guides the assembly to a safe impact area.

Although the technique cited above offers many advantages, it is comparatively new. The choice of this technique and its design parameters was largely based on work accomplished by United Technology Center under contract with the Air Force Rocket Propulsion Laboratory. In this program, the general operating characteristics of aft-end pyrogen ignition were first studied in depth. The conclusions made in the final report (1) on this contractual effort were:

a. The relationship between igniter mass flow rate and motor ignition delay for aft-end mounted pyrogens is quite similar to that of the head-end mounted pyrogen igniter used previously in many solid rocket systems.

b. It appeared that the igniter gases did not penetrate to the head end of the motor being ignited, but formed a stagnation plane within the port cavity.

c. If the clearance area between the igniter hardware and the main motor exit cone (A_c) is greater than the throat area (A_t) of

the motor being ignited, there is no possibility of a main motor ignition overpressure. To characterize this positioning parameter, United Technology used the term ϵ^* which is equal to A_c/A_t.

The validity of the first conclusion was amply demonstrated by a number of motor firings in which motor size and aft-end pyrogen mass flow rate were varied over a wide range. The tentative conclusion regarding igniter penetration was based on an interpretation of the relative response times of thin film thermocouples which were distributed along the port on the propellant surface. The data showed a rapid rise for thermocouples mounted at the aft-end of the motor grain, (presumably signifying contact with igniter gases and subsequent propellant ignition). Following a delay period, the forward mounted thermocouples then responded in sequence along the grain. The effect of ϵ^* on motor pressure rise time and overpressure potential was emperically determined. The mechanism of the modulation in pressure rise rate due to variation of ϵ^* was not established. The design relationship for no overpressure (viz. $\epsilon^* \geq 1$), was based on a series of firings where ϵ^* was systematically varied between 0.8 and 1.8. Only in the case $\epsilon^* = 0.8$ did an overpressure occur. It was then hypothesized that the mechanism for the overpressure was simply physical throat blockage, i.e., for $\epsilon^* < 1.0$ the throat area is smaller than the design area and the motor consequently operates at a higher than normal pressure. Further, for $\epsilon^* > 1.0$, sufficient flow area is available for normal motor operation

and the flow from the main motor "shuts off" the igniter gas flow into the motor as soon as sonic conditions are established in the motor throat. With the motor thus isolated from the igniter flow, no perturbation of the motor operating pressure was presumed possible.

The Air Force Large Solid Rocket Motor Programs were initiated a short time after the publication of these conclusions. All but two of the large motor firings are to use aft-end pyrogenic ignition. With the exception of an abbreviated cold flow study to be run at Aerojet-General Corporation to verify a new theoretical approach to prediction of penetration depth (2), there was general acceptance of the United Technology Center results. No extensive ignition characterization programs appeared justified.

Two of the conclusions, however, did not appear solidly founded and were the subject of this investigation. The existence of the stagnation plane within the motor port had been observed only by inference. The second area for study concerns the conclusion stating the impossibility of motor overpressure with aft-end pyrogenic ignition. If the latter conclusion were incorrect, a potential hazard to the successful firing of the large solid motors would exist. Secondly, if the possibility of overpressure was real, this fact would have implications in the planning for necessary pyrogen release systems or other protective mechanisms in the proposed facilities for the large solid motor launch sites.

Based on these questions regarding the validity of United Technology Center's aft-end pyrogen performance predictions, an in-house program, the Solid Propellant Rocket Ignition Test and Evaluation Program (SPRITE), was initiated at the Air Force Rocket Propulsion Laboratory. Its objectives were: (1) to demonstrate the characteristics of the aft-end pyrogen stagnation zone; and (2) to determine the possibility or impossibility of aft-end pyrogen generated motor throat blockage and consequent motor overpressure. A third objective, predicated on the existence of ignition overpressures, was the determination of methods to eliminate this hazard by variation of igniter operating characteristics. The results of these studies constitute the substance of this paper.

PART I

AFT-END PYROGEN PENETRATION STUDY

The objective of this study was to qualitatively determine the existence or non-existence of a stagnation plane within the port of a solid rocket motor under the influence of the discharge of an aft-end pyrogen igniter. The importance of this validation rests in the fact that the character of motor ignition depends upon the fraction of the propellant grain which is ignited by the action of the igniter. In theory, the ignition of even a small portion of the propellant surface will accomplish ignition of the motor. In this case, however, the ignition process is slow, frequently associated with high ignition pressure peaks and is generally undesirable.

The phenomena of aft-end pyrogenic ignition is presumed to follow the following sequence:

1. The pyrogen is ignited and discharges a jet through the nozzle of the motor being ignited.

2. The cross-section of the igniter jet increases until it reaches the port walls and then, acting as a piston, compresses the gas in the head end of the motor.

3. Equilibrium is established when flow out of the motor balances the igniter jet mass flow rate. At this time, a stable stagnation plane is formed within the motor port. The position of this plane is stable as long as no propellant ignition takes place and as long as the igniter mass flow rate remains constant.

4. The igniter gases bathe that fraction of the propellant surface aft of the stagnation zone. In this area, the ignition process is similar to that of a head-end mounted pyrogen igniter.

5. The ignited boundary on the propellant surface then moves toward the head end of the motor under the influence of radiative flux from the body of hot gases contained in the aft portion of the port.

This ignition model reveals the importance of the existence of a stagnation zone in determining the character of aft-end pyrogenic ignition. Since a mechanism is presented which limits the fraction of propellant surface that is exposed to igniter thermal flux, the position of the stagnation plane will have a profound effect on the

characteristic ignition time. High penetrations, (70-80%), would give rapid ignition having characteristic times comparable to those of head-end mounted pyrogenic ignition. Low penetrations would yield longer ignition transients with slow pressure rise rates since the propagation rate of the ignition front into the stagnant head end region is very slow as compared with ignition delays for propellant directly exposed to igniter exhaust gases.

A second important characteristic of the stagnation zone is its stability. If the zone forms but is unstable in time, its importance in controlling the ignition transient is correspondingly reduced. If, for example, the aerodynamic equilibrium is only briefly established and then rapidly breaks down and permits flow of hot igniter exhaust into the previously isolated head end, ignition will be more rapid and the ignition delay will approach that of the head-end pyrogenic igniter.

THEORETICAL ANALYSIS

An analytical model for prediction of equilibrium penetration has been developed by Plumley. The analysis is given in detail in Reference 2 and will only be summarized here. The model is based on the assumption of equilibrium such that the mass flow in through the nozzle throat equals the mass flow out; i.e., propellant ignition is not permitted. It is assumed that the outward flow from the motor is choked at the nozzle throat and that the incoming igniter gas has expanded to the choked flow static pressure. The primary igniter

gas flow thus uses a portion of the motor throat area and the returning gas passes through the remaining annular space at the same static pressure and at sonic velocity. This continuity determines the equilibrium pressure level in the motor throat.

The second basic assumption is that the incoming flow is turned around in order to flow out of the motor. Therefore, using a momentum balance, the pressure in the head end of the motor is determined. Other assumptions include: one dimensional flow, no mixing between igniter and cold gases, isentropic compression of the trapped gases, no mixing between incoming and outgoing gas, and that the perfect gas laws apply.

Formulation of the equations appropriate in defining mass and momentum conservation and boundary conditions yields a system of equations requiring computer solution. The results are presented graphically in Figure 2. It should be noted that the theoretical results do not predict an effect of ℓ^* or any other igniter positioning parameter. The only non-geometric parameters are the igniter chamber pressure, (constant and equal to 1000 psia in Figure 2), and the total igniter mass flow rate.

APPARATUS

In order to rapidly evaluate the penetration phenomena, a simple technique was chosen. The apparatus used is shown in Figure 3. The propellant grain was designed to provide a constant surface area, (and

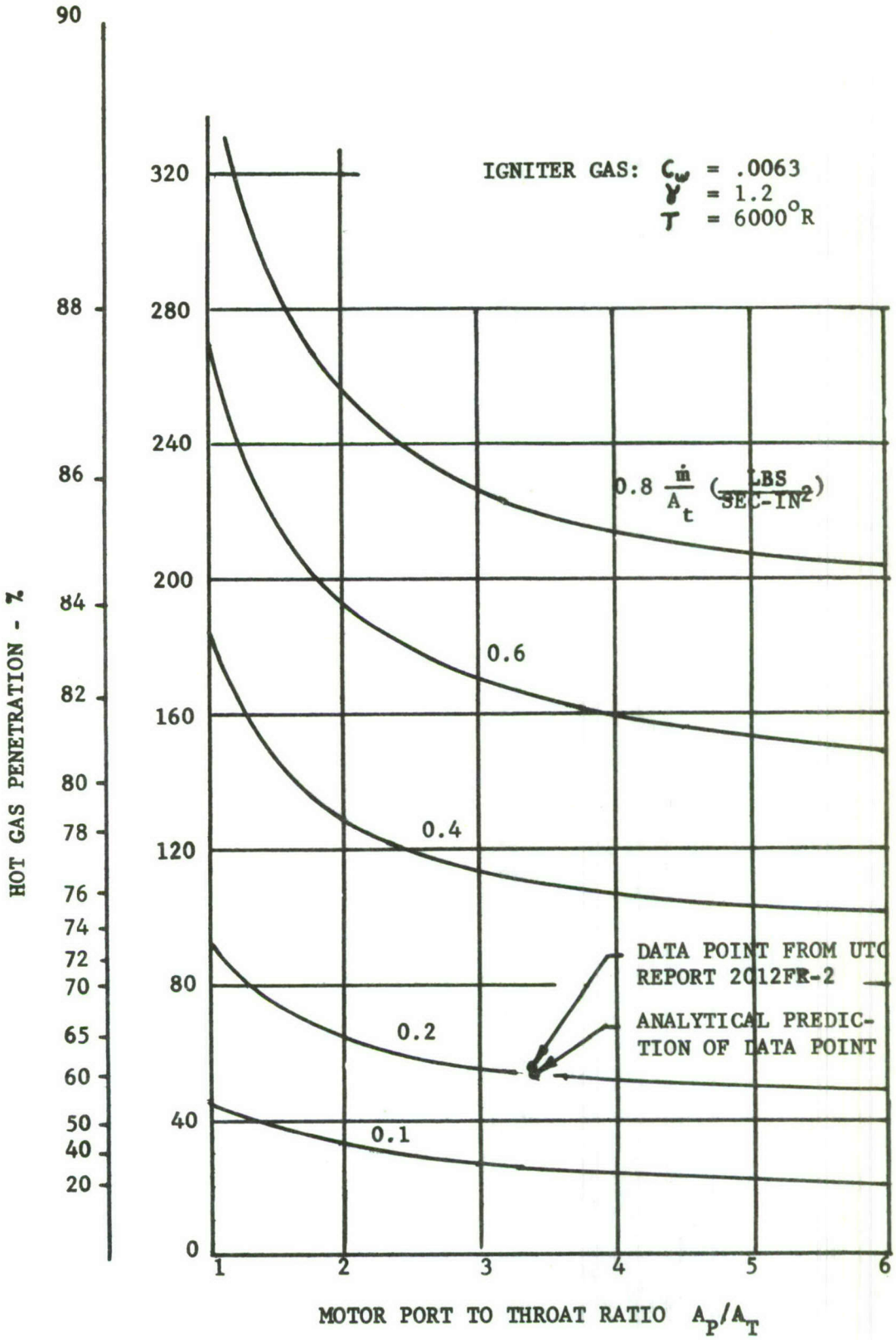


FIGURE 2 - THEORETICAL AFT-END PYROGEN PENETRATION CORRELATION (AFTER PLUMLEY - REF. 2)

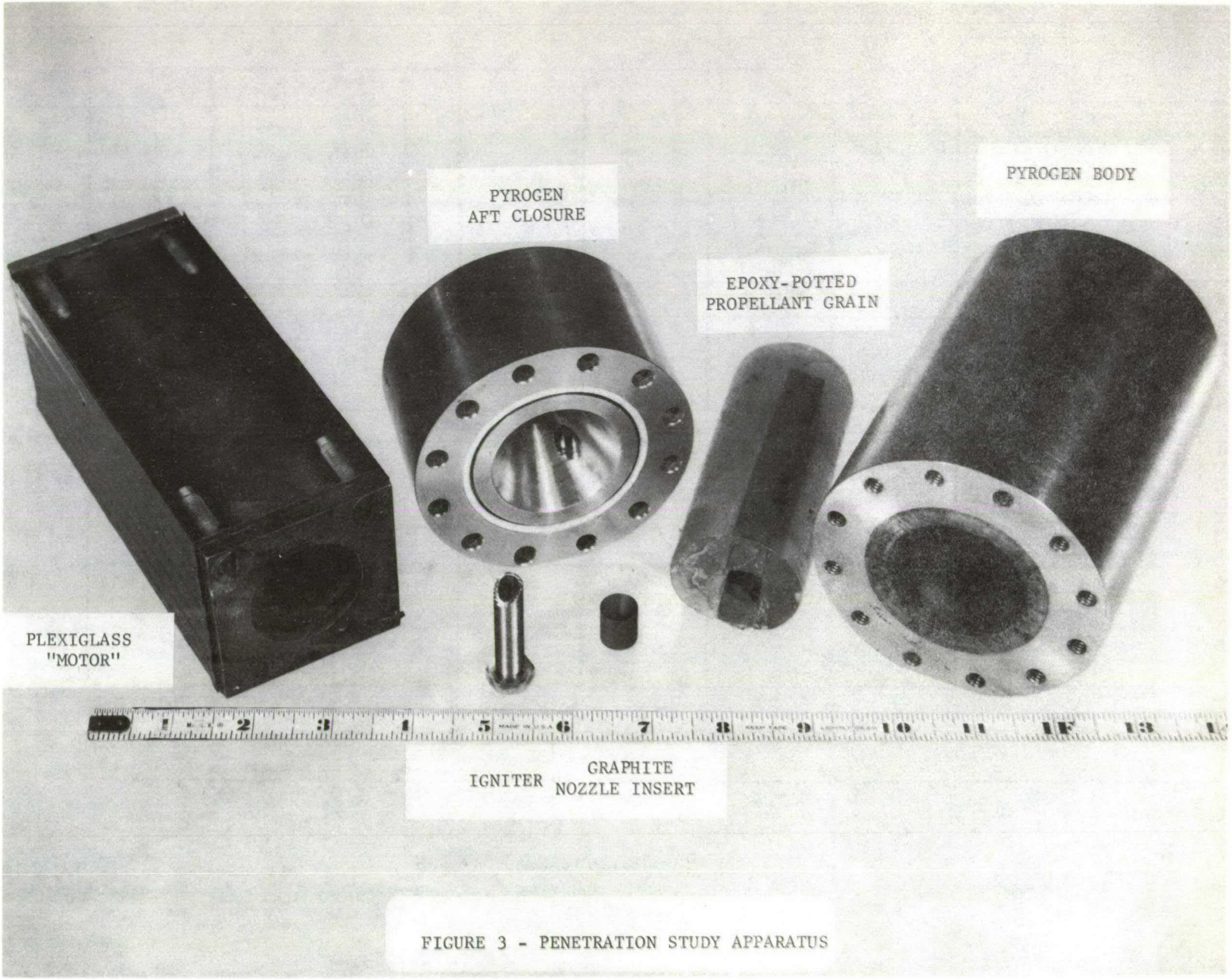


FIGURE 3 - PENETRATION STUDY APPARATUS

consequent motor pressure and mass flow rate), during the burning time. The design operating pressure was 1000 psi. The nozzle throat diameter was .125 inches.

The simulated motor was made of heavyweight Plexiglass containing a cylindrical cavity. A metal plate with gasket seal was fastened to either end. One of the plates was perforated to simulate the motor throat.

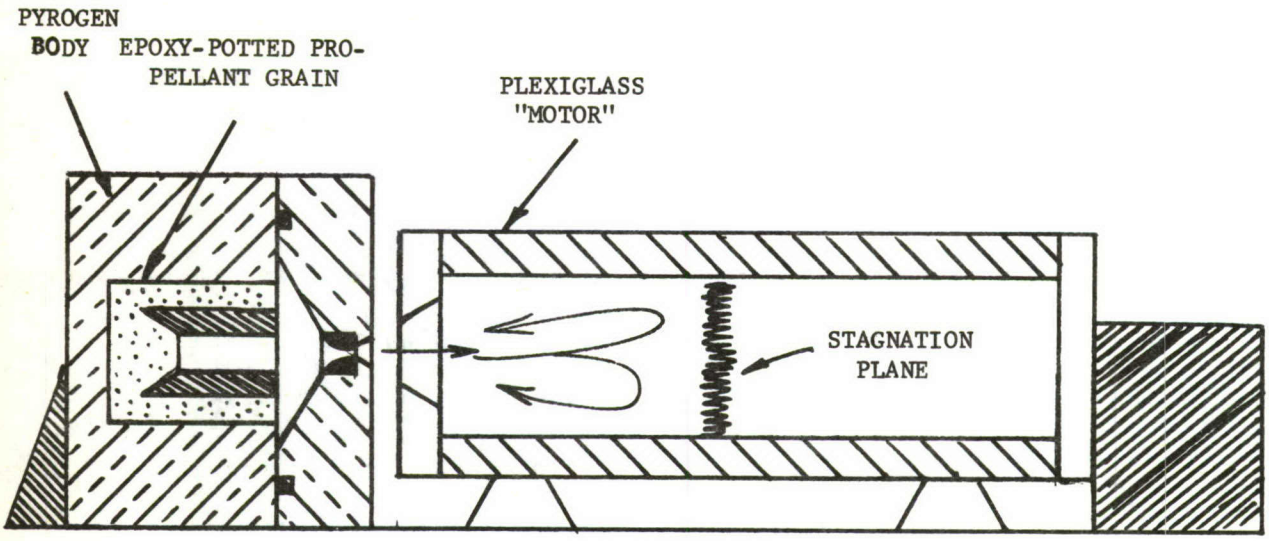
The experimental arrangement for a typical run is shown in Figure 4. The c^* parameter was simulated by control of the spacing between the exit plane of the pyrogen expansion cone and the face of the "motor" nozzle plate.

Data acquisition was accomplished with a 500 frames/second 35mm motion picture camera. Using high-speed black and white film, the penetration phenomena was thus recorded at 2 msec intervals.

RESULTS

A total of 11 firings were conducted. Of these, only four gave meaningful results. In the other firings, malfunction or obscurement of the field of view by pyrogen exhaust gases rendered the data unusable.

The data are presented in Figures 5 through 9. The operating conditions, data on "motor" configurations and penetration results are presented in Table I.



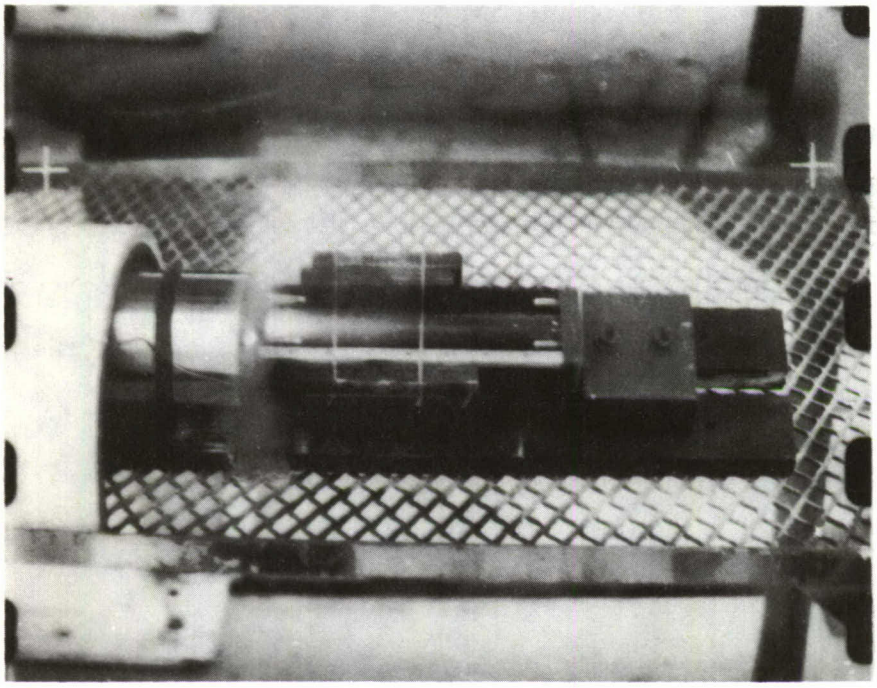
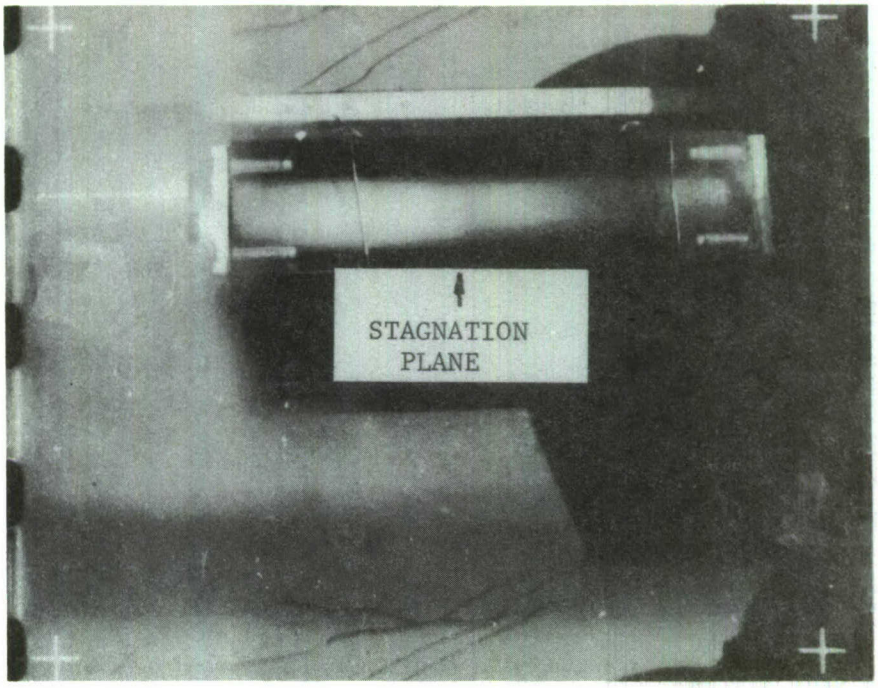


FIGURE 4

EXPERIMENTAL CONFIGURATION - PENETRATION STUDY



RELATIVE TIME = 0 (ABOVE)
RELATIVE TIME = 32 msec (BELOW)

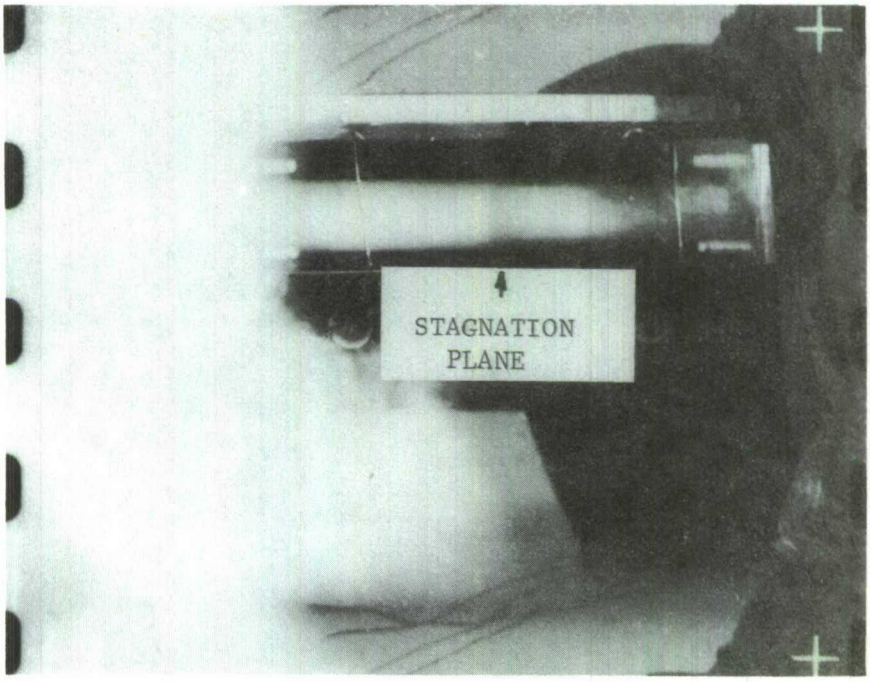
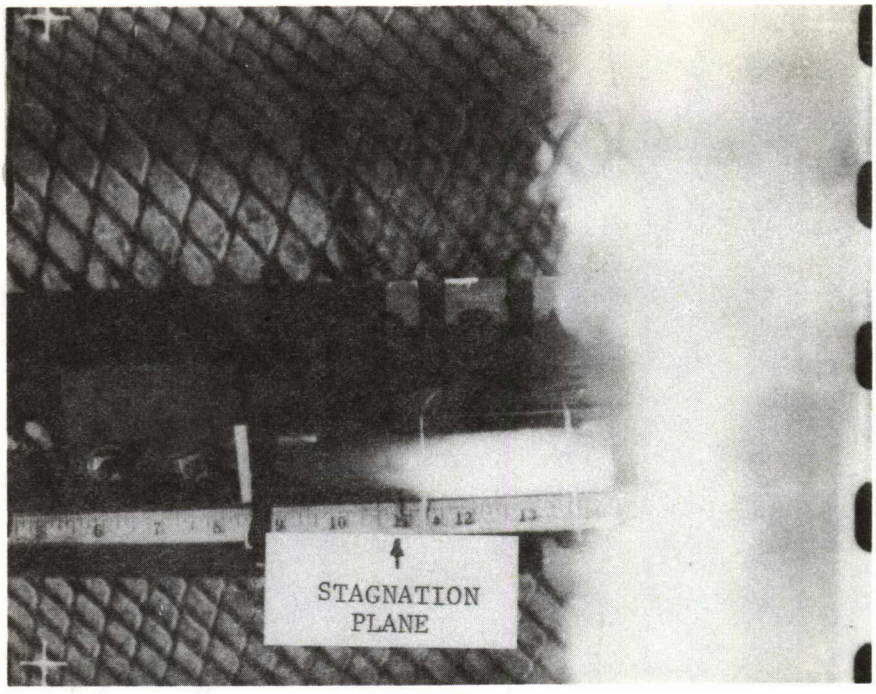


FIGURE 5

DATA OF RUN 2



RELATIVE TIME = 0 (ABOVE)

RELATIVE TIME = 174 msec (BELOW)

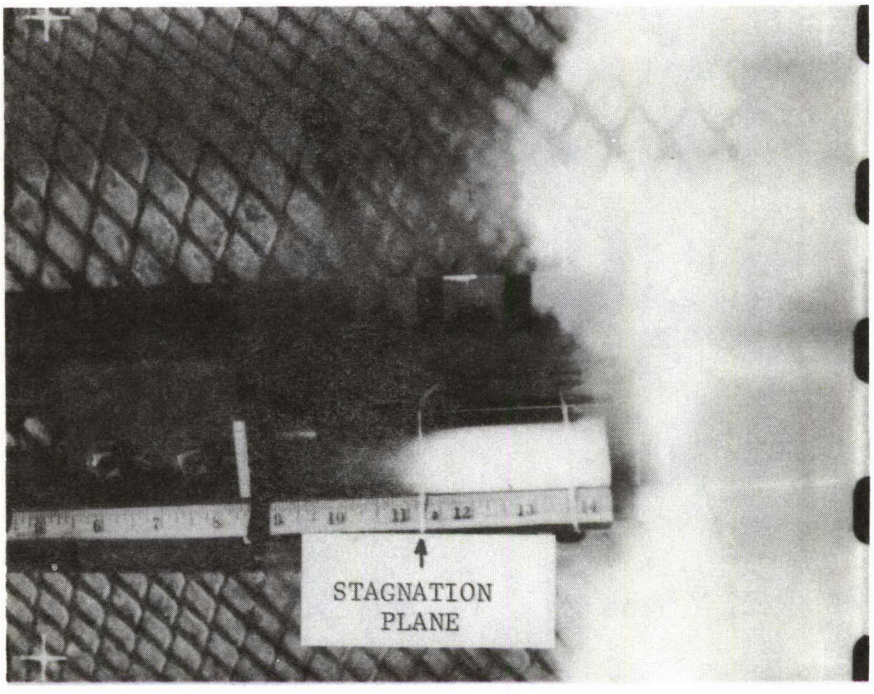
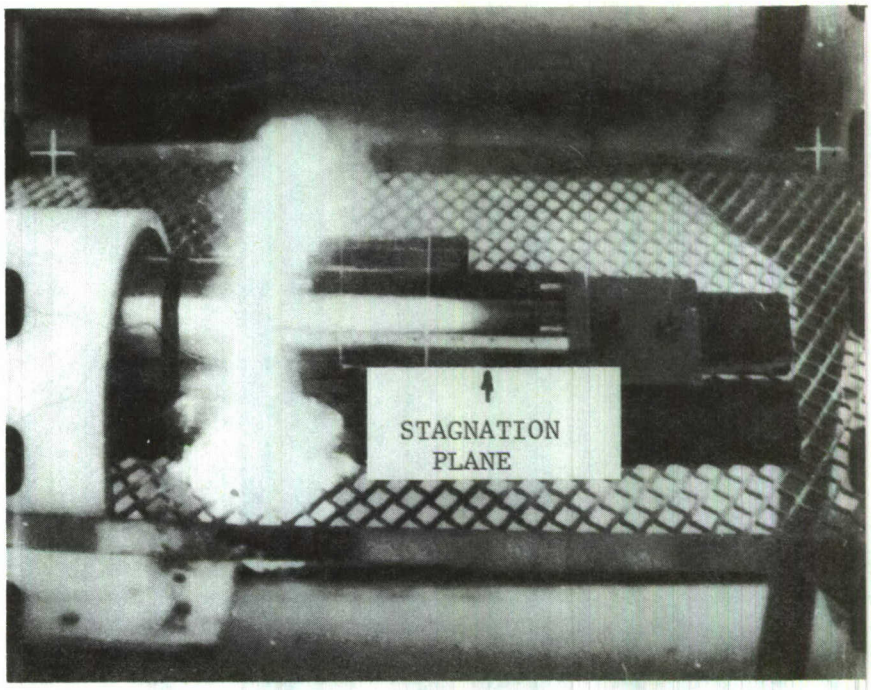


FIGURE 6

DATA OF RUN 3



RELATIVE TIME = 0 (ABOVE)

RELATIVE TIME = 8 msec (BELOW)

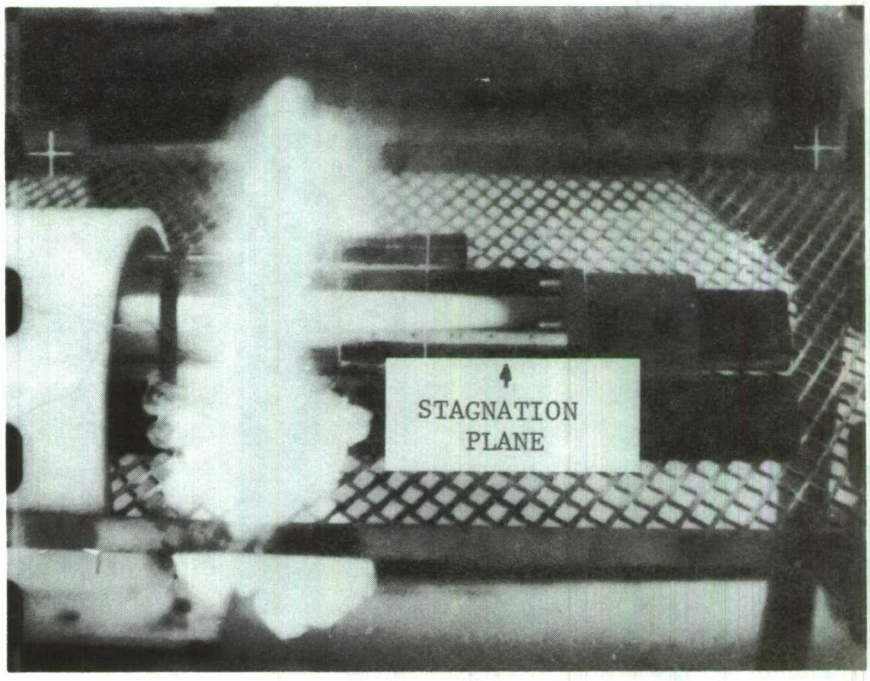


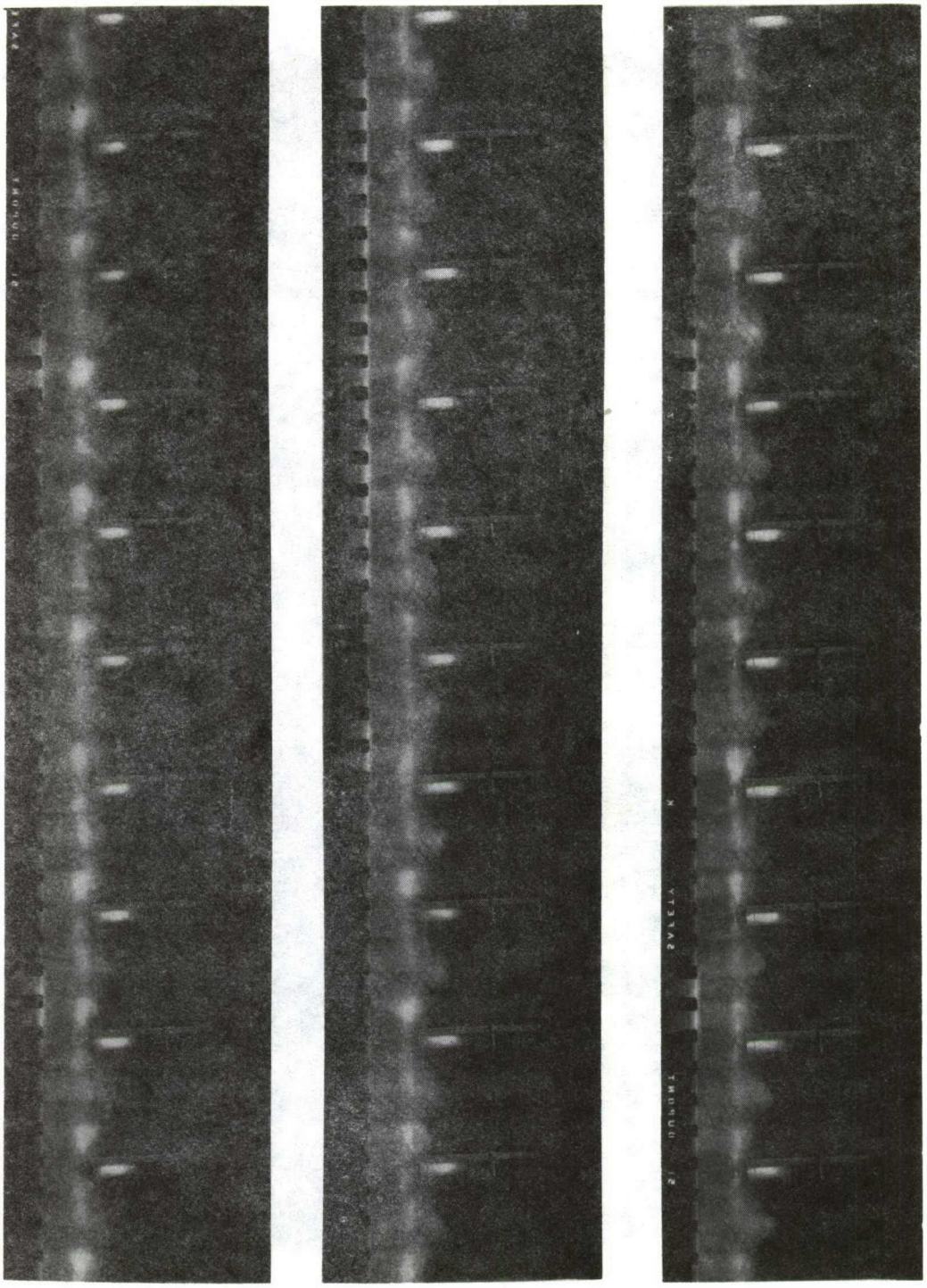
FIGURE 7

DATA OF RUN 7

TIME = 0

TIME = 256 msec

TIME = 392 msec



TIME = 18 msec

TIME = 274 msec

TIME = 400 msec

FIGURE 8

SEQUENCE PHOTOGRAPHS - RUN 3

TABLE I

A. DATA TABULATION

Run	D_p	D_t	A_p/A_t	*	$\frac{\dot{m}_i}{A_t}$	L	L/D_p	% Penetration Predicted	% Penetration Observed
2	1.25	0.895	1.3	1.3	.125	9	7.2	56	50 ± 5
3	1.25	0.895	1.3	1.3	.125	6	4.7	56	48 ± 4
7	0.75	0.655	1.3	2.0	.233	8	10.7	73	68 ± 5
11	1.75	1.535	1.3	1.3	.042	4	2.3	Small	100

B. PYROGEN MOTOR MASS FLOW RATE

$$\dot{m}_i = P_i C_w A_t$$

$$= (1000)(0.0064)(\ \ /4)(0.125)^2$$

$$= 0.0786 \text{ lb/sec}$$

C. NOMENCLATURE

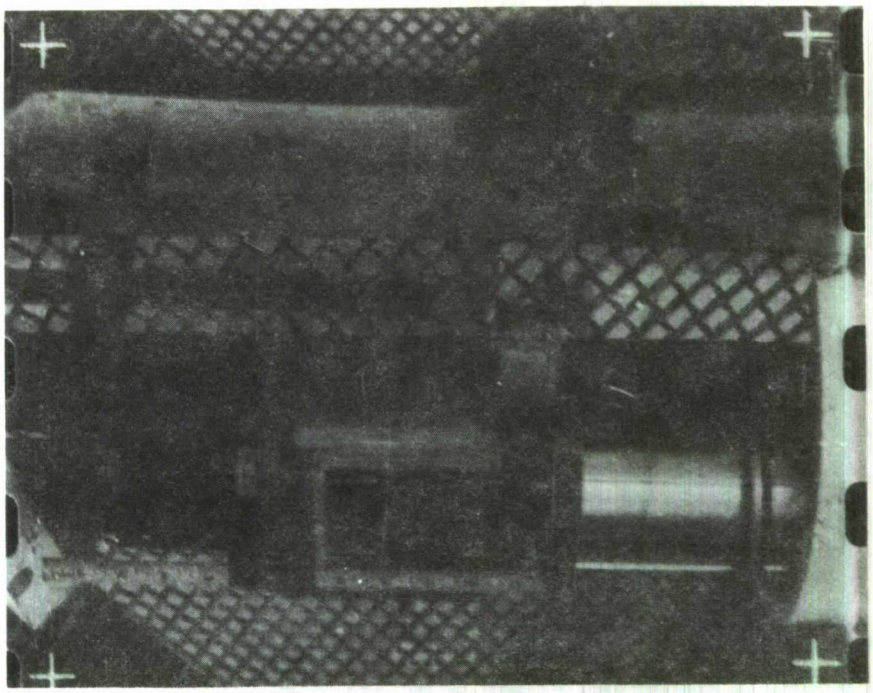
D_p	Port Diameter	(in)
D_t	Throat Diameter	(in)
A_p	Port Area	(in^2)
A_t	Throat Area	(in^2)
L	Motor Length	(in)
$\dot{m}_i$	Igniter Mass Flow Rate	(lb/sec)
P_i	Igniter (Pyrogen) Chamber Pressure	(psia)
C_w	Flow Coefficient	(sec^{-1})

DISCUSSION OF RESULTS

The photographs of the stagnation phenomenon, Figures 5 through 7, are remarkable. They are representative of all pictures taken during these runs before the view became obscured with smoke. The existence of a well defined stagnation zone is clearly shown in all cases. The appearance of low density smoke in the head end is mainly due to a small amount of burning igniter material projected into the "motor" during the ignition phase of the pyrogen unit.

In all cases, the position of the stagnation zone was remarkably stable. The contact strip photos in Figure 9 show the positional stability of the stagnation plane over a period of 400 msec. This period is long as compared to typical solid propellant rocket motor ignition delays.

The agreement between theory and experiment in defining the position of the stagnation zone is similarly striking (Table I). The experimental values are within 10% of the predicted penetration depths. Although some runs showed a slight movement of the stagnation plane towards the head closure during the run, this is most likely due to a slow increase in the igniter mass flow rate as the run progressed. Because of the small size of the pyrogen throat, deposition of aluminum in the propellant combustion products tended to restrict the throat as the firing proceeded, thus increasing igniter chamber pressure and mass flow rate. This constriction was observed in post-firing inspection of the nozzle assembly.



RELATIVE TIME = 0 (ABOVE)

RELATIVE TIME = 2 msec (BELOW)

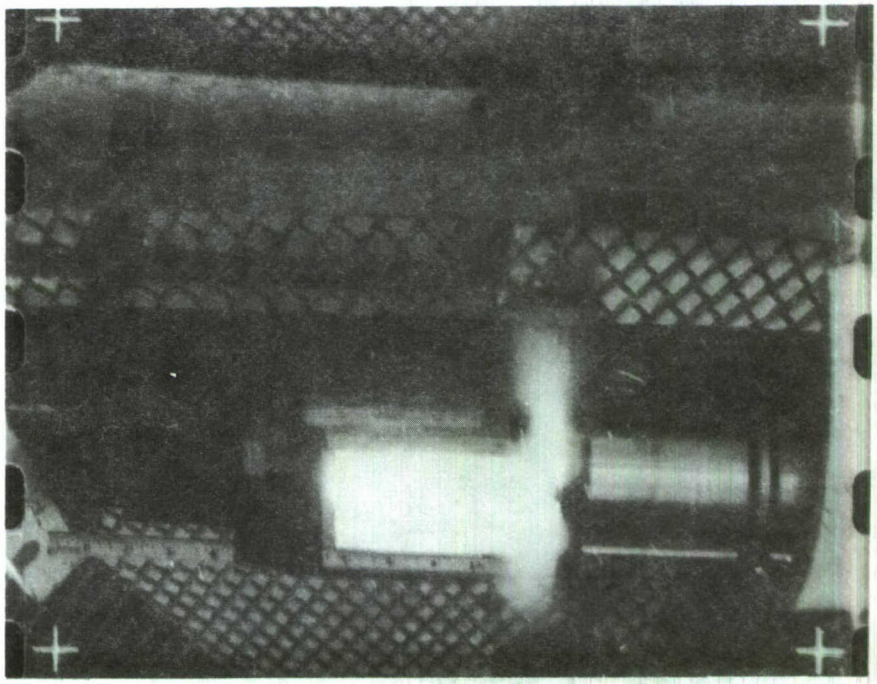


FIGURE 9

DATA OF RUN 11

The effect of ϵ^* appears either small or beyond the resolution of the data. It must be noted, however, that the ϵ^* simulation in this experiment is somewhat superficial and one may observe some effect in the "real" case.

The cold-flow experiments of Plumley (2) which followed this work showed similar agreement to theory, including a lack of marked dependency of penetration depth on ϵ^*.

In Run 11, there was a marked deviation from theory. In this case (Figure 9), the pyrogen gases clearly penetrated to the head closure. The explanation for this is clear. In order for the analytical model to be valid, the igniter jet boundary must contact the port walls. Only after this takes place can a true "piston" effect be realized. Taking a 15° half-angle as a growth rate typical of free jet diameters, the piston "sealing" would not take place until the jet was about 1/2 inch from the head closure plate. For this reason, unique to this run, a stagnation zone was never formed. This effect should be considered when motors of low length to diameter ratios are being analyzed in the design of aft-end pyrogenic igniters.

CONCLUSIONS

From the results of this program, the following conclusions can be made:

1. In the operation of aft-end pyrogen igniters, a definite and relatively sharp-edged stagnation plane is formed within the port of the motor being ignited.

2. This stagnation zone is position-time stable as long as pyrogen chamber pressure is constant.

3. The theoretical analysis of Plumley appears valid within experimental error.

4. The effect of ϵ^* appears small in defining penetration depth of the stagnation plane.

5. If the length to diameter ratio of the motor port is small, penetrations may occur which are greater than predicted by theory.

PART II

AFT-END PYROGEN THROAT BLOCKAGE STUDY

The objective of this phase of Project SPRITE was to study the throat blocking effect of an aft-end pyrogen on a solid rocket motor when both were operating at full chamber pressure. This study was initiated when review of the UTC report (1) and the films of the test firings revealed that the pyrogen was always expelled from the nozzle region before the motor chamber pressure reached 85 percent of normal operating pressure. Based on this evidence, it was believed that the conclusion stated by UTC in this report regarding the impossibility of motor overpressure when using an aft-end pyrogen igniter at $\epsilon^* \geq 1.1$ had been formulated without wholly definitive data. UTC had not correlated the time phasing of motor chamber pressure with the pyrogen release.

The aft-end pyrogen ignition systems planned for 156 inch and 260 inch motors are designed for use with a release mechanism. The

release mechanism relies on the proper operation of explosive bolts which are actuated when the motor chamber pressure reaches approximately 20 percent of normal operating pressure. Failure of the pressure sensor or bolt actuation systems can cause the pyrogen to remain in position when both the pyrogen and the motor are at full operating chamber pressure. It should also be noted that in normal launch configurations using an aft end pyrogen igniter, no release mechanisms are planned so that this same situation can develop. This phase of the SPRITE study was devoted to determining the extent of the motor overpressure hazard that may be associated with this situation.

THEORETICAL ANALYSIS

The phenomena under investigation may be considered as a dynamic, non-equilibrium flow between two opposed supersonic jets confined within rigid boundaries. A major effort was expended by the authors in order to obtain a mathematical model of this phenomena. This model was unobtainable due to the extreme complexity of highly turbulent, two-phase, non-steady, non-isentropic mixing flows. Consequently, only a qualitative analysis was made.

When the motor reaches normal operating pressure with the pyrogen operating, the condition exists where two supersonic gas streams are flowing directly against each other. The penetration of the pyrogen flow across the throat plane of the motor will reduce the effective motor throat area. The effect of this reduced area is to

raise the pressure inside the motor. The magnitude of this pressure increase depends upon the propellant burning rate exponent, n. This latter point will be discussed in more detail in the section on discussion of results.

Assuming the velocity of the pyrogen remains constant, the momentum of the pyrogen flow will be a function of the pyrogen chamber pressure. The greater the pyrogen pressure the greater the flow momentum and, hence, the penetration of the flow into the motor. It is postulated that the interchange of momentum between the two streams is statistical. This implies that the amount of momentum per unit area and the stay time of the fluid particles are of prime importance. The stay time is defined as the time for a particle of pyrogen fluid to travel from the pyrogen exit cone to the motor throat. Assuming the pyrogen exhaust is cylindrical in shape and both the pyrogen and motor are operating, the momentum per unit area is a function of the respective chamber pressures only. If these pressures are constant, then the interchange of momentum will vary only with stay time. Thus, with constant pyrogen exhaust velocity, as c^* decreases, the probability of penetration increases. Also, if the motor pressure is held constant and the pyrogen pressure is allowed to increase, then at constant stay time (c^* = constant) the penetration and effective throat blockage would increase.

This description of the penetration behavior of aft-end pyrogen units implies the existence of a degradative mechanism and that the

effects of this mechanism, (reduction in penetration level), are felt more strongly as the distance between the pyrogen and motor throat plane is increased. This degradation is suggested in Plumley's data (2). His data shows a consistent, but small, reduction in penetration level with increasing ϵ^*. The low magnitude of the ϵ^* effect in Plumley's work is presumably due to the low relative counter mass flow as compared to that in operating motors, (Plumley, 1:1 versus typical motors, 20:1).

A further speculation concerns the validity of retaining ϵ^* as the only pertinent positional parameter defining penetration. Since the phenomenon is closely tied to the behavior of jets, one should consider a parameter more closely associated with the pyrogen, (or motor), jet configuration. Such a parameter might be the ratio of the distance between the pyrogen throat and the motor throat plane to the pyrogen throat diameter (x_o/D_t). Since a single geometrical configuration was used in this experimental program, differentiation of the relative importance between ϵ^* and x_o/D_t could not be made. For consistency with accepted aft-end pyrogen technology, the data from this effort are reported and correlated in terms of ϵ^* only.

APPARATUS

The test apparatus consisted of an M-8 JATO unit and an Air Force Rocket Propulsion Laboratory designed pyrogen. The pyrogen was rigidly attached to the JATO exit cone by four bolts. The distance between pyrogen and motor throat was varied by use of machined

cylindrical spacers. The assembled test unit is shown in Figure 10. The minimum exhaust area of this test configuration is 27 square inches. The throat area of the JATO is 1.33 square inches. Other dimensional characteristics of the test apparatus are given in Table II. The disassembled pyrogen and JATO nozzlesare shown in Figure 11.

It is noted in Figure 12 that the pressure time trace of the JATO immediately after ignition is quite irregular. Following this transient period, the pressure remains relatively constant. The ignition spike and the ensuing saddle characteristics of the M-8 JATO required a delayed ignition of the pyrogen in order to obtain data from a stable reference motor pressure. The JATO is ignited by a conventional, head-end JATO igniter. The pyrogen pressure-time trace, reference Figure 13, was designed to be progressive. This feature allows the determination of the effect of increasing pyrogen mass flow or momentum in a given test. Pyrogen and JATO pressures were determined with Statham and Tabor pressure transducers, respectively, and recorded on magnetic tape.

RESULTS

A total of seven runs were successfully completed. Four of these tests were conducted at $\epsilon^* = 1.2$, one at $\epsilon^* = 1.4$ and two at $\epsilon^* = 2.0$. These data are presented in Figure 14. In all firings at $\epsilon^* = 1.2$ and 1.4, JATO overpressurization was evident. At $\epsilon^* = 2.0$ there was no evidence of throat penetration or overpressure. The data

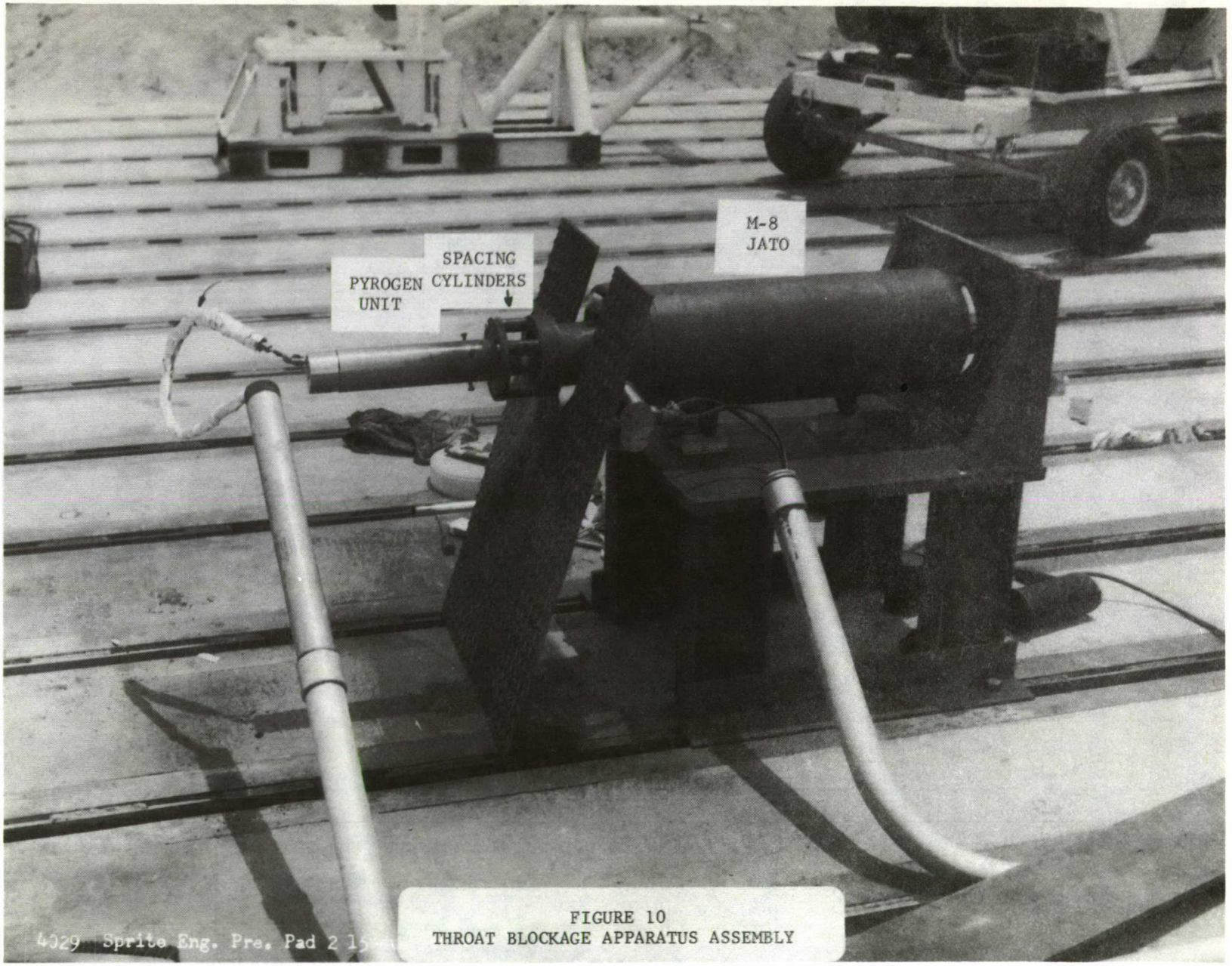


FIGURE 10
THROAT BLOCKAGE APPARATUS ASSEMBLY

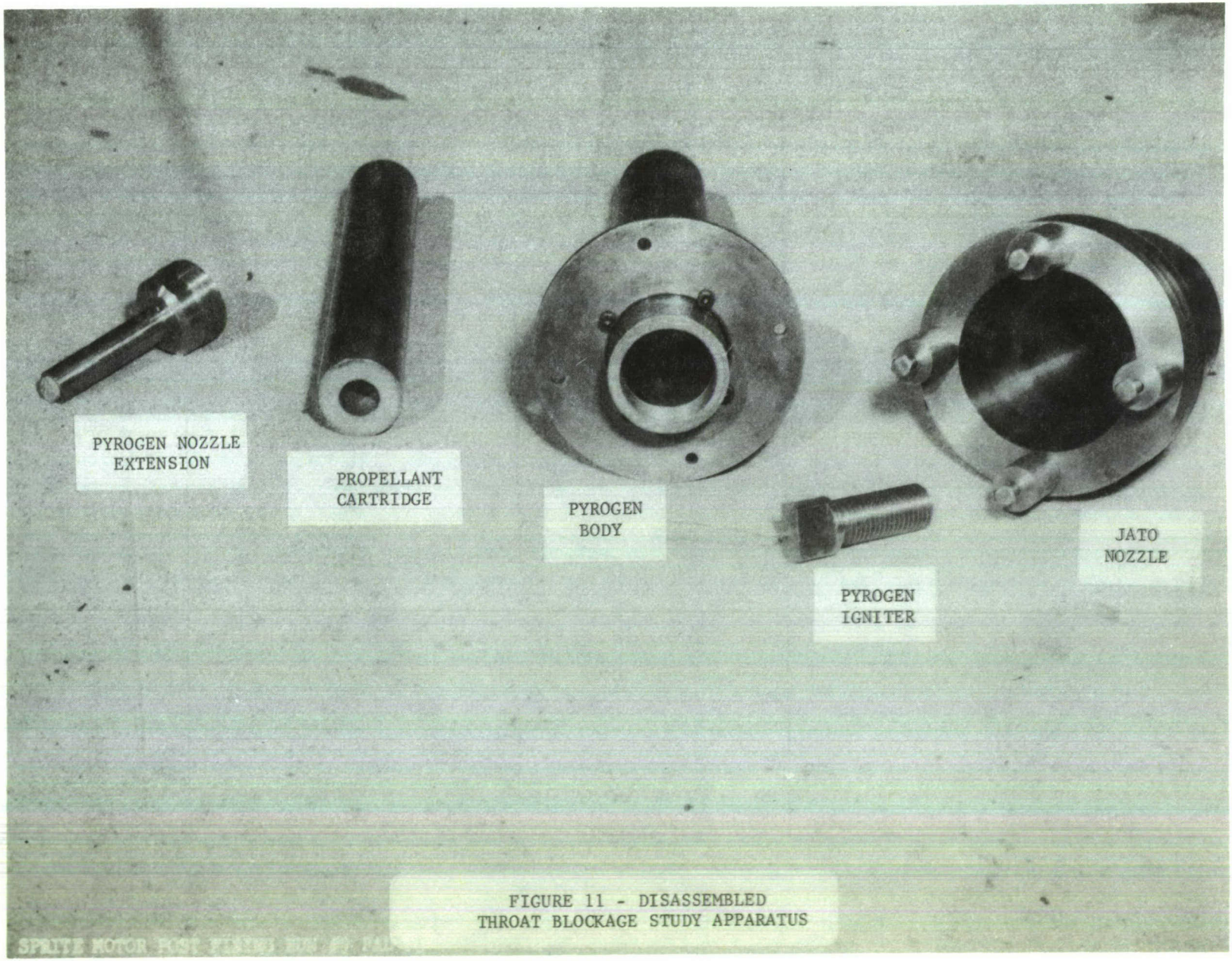


FIGURE 11 - DISASSEMBLED
THROAT BLOCKAGE STUDY APPARATUS

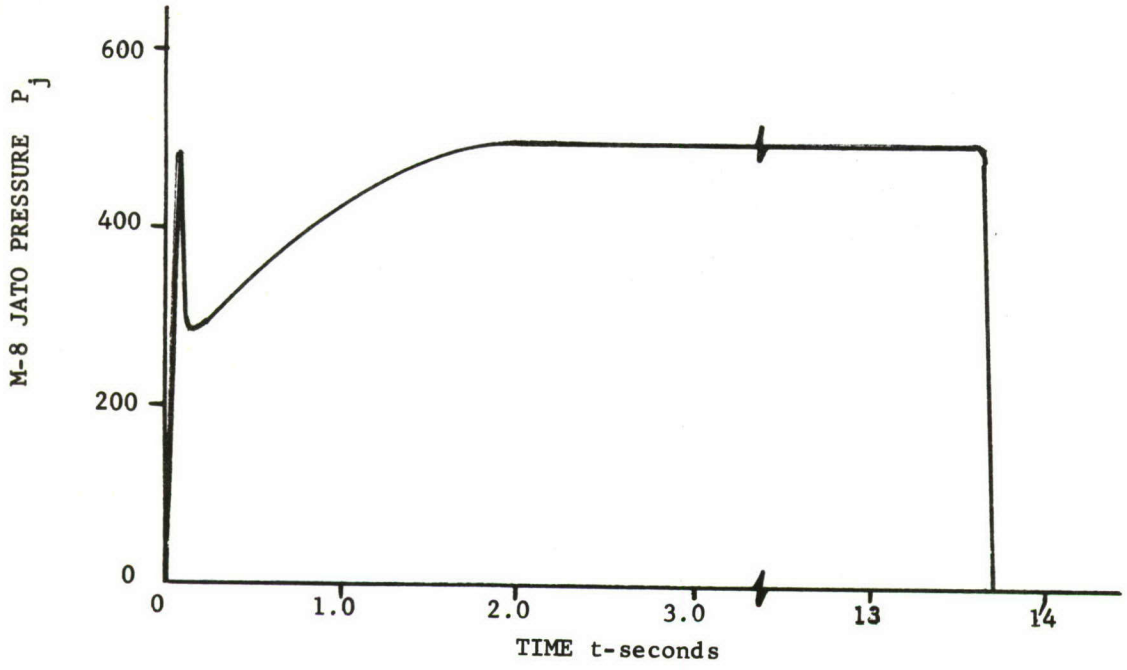


FIGURE 12 - TYPICAL M-8 JATO PRESSURE-TIME TRACE WITHOUT AFT-END PYROGEN FIRING

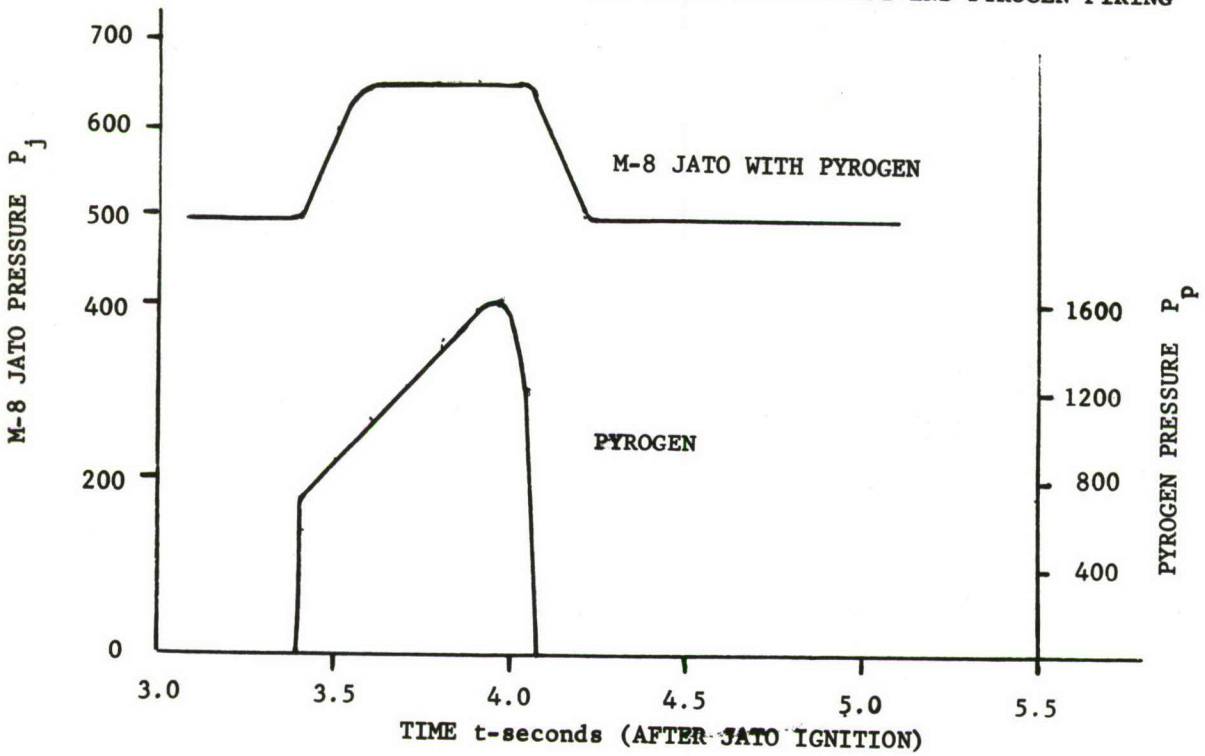


FIGURE 13 - TYPICAL SEGMENT OF JATO PRESSURE TIME TRACE, $\epsilon^* = 1.2$, WITH PYROGEN FIRING; PYROGEN PRESSURE TIME TRACE.

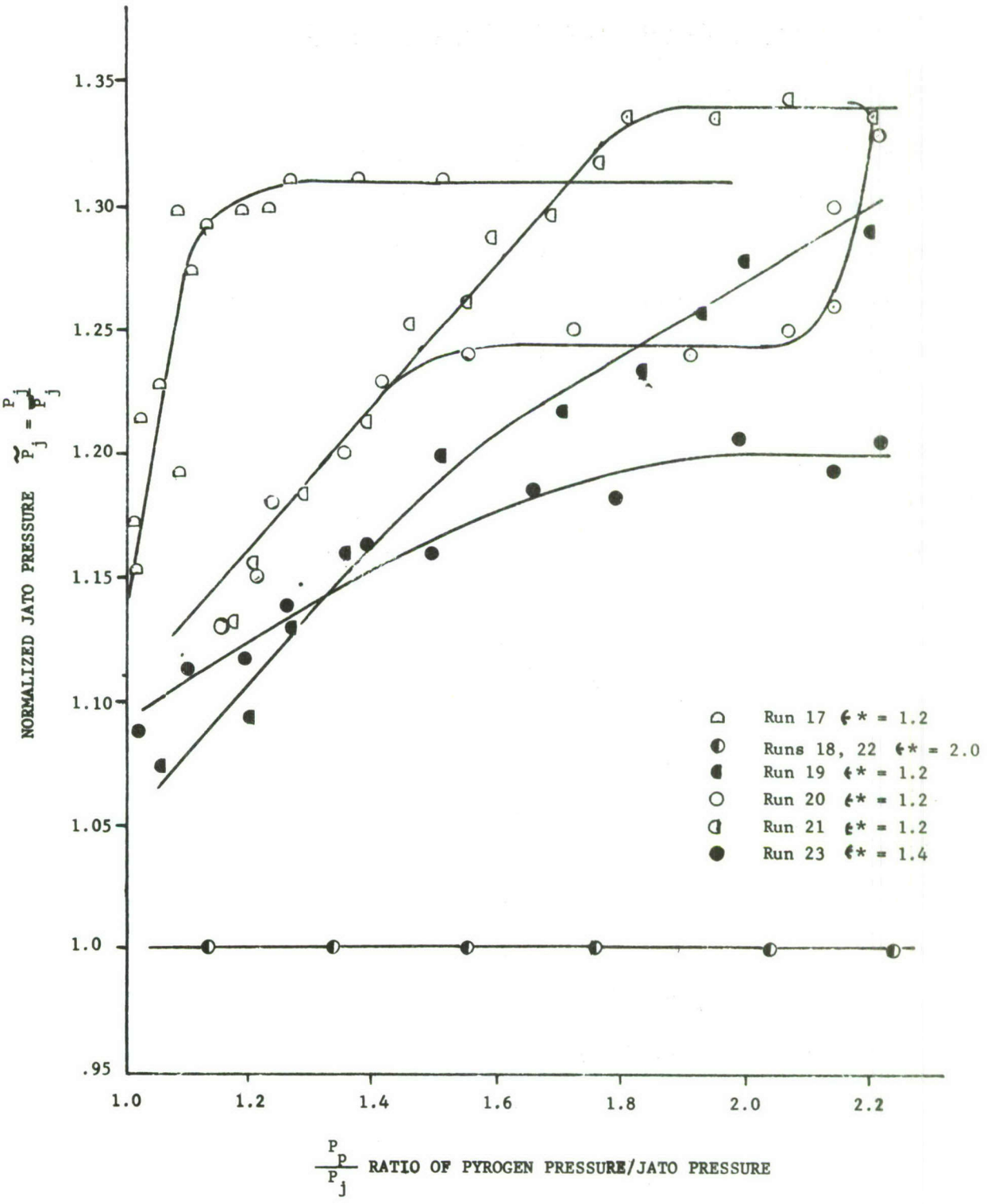


FIGURE 14 - NORMALIZED TEST DATA FOR AFT-END PYROGEN/M-8 JATO TESTS

TABLE II

APPARATUS GEOMETRICAL AND OPERATIONAL CONSTANTS

A. GEOMETRICAL CONSTANTS

Item	Units	Symbol	Pyrogen	Motor
Throat Diameter	in	D_t	0.342	1.304
Exit Plane Diameter	in	D_x	1.004	3.845
Maximum Hardware Diameter	in	D_{max}	1.200	N/A
Throat Area	in^2	A_t	0.0918	1.336
Exit Plane Area	in^2	A_x	0.798	11.630
Nozzle Half-Angle	deg	θ	15°	15°

B. CALCULATED CONSTANTS

Distance (x_o) between Pyrogen Exit Plane and Motor Throat Diameter, (θ and D_t refer to motor parameters).

$$x_o = \frac{D_o - D_t}{2 \tan \theta} + \frac{(D_o - D_{max}) \tan \theta}{2}$$

$$D_\epsilon^2 = D_{max}^2 + \epsilon^* D_t^2 \cos \theta$$

ϵ^*	x_o (in)	x_o/D_t(igniter)
1.2	1.097	3.21
1.4	1.272	3.72
2	1.762	5.16

C. OPERATIONAL CONSTANTS

1. Mass Flow Rates (lb/sec)

Pressure	JATO $\dot{m}$	Pyrogen $\dot{m}$
500	5.0	0.293
550	5.5	0.320
600	6.0	0.350

1. Mass Flow Rates (lb/sec) Continued:

Pressure	JATO $\dot{m}$	Pyrogen $\dot{m}$
650	6.5	0.378
800	8.0	0.466
1000	10.0	0.583
1200	12.0	0.700
1400	14.0	0.816

2. Propellant Characteristics

	JATO	Pyrogen
Propellant Designation	OGK	Similar to Rohm & Haas C112cb
Propellant type	double-base	Aluminized-composite double-base
Burning rate pressure exponent	0.0	0.45
Flow Coefficient (sec^{-1})	.00747	.00635
Specific heat ratio	1.23	1.19

TABLE III

REDUCED EXPERIMENTAL DATA

Run	ϵ*	Pyrogen P_{pi}	Pyrogen P_{pf}	JATO $P_{j\text{-ave}}$	JATO $P_{j\text{-max}}$	$D_{B_{max}}$	$\tilde{P}_{j\text{-max}}$	$\frac{dP_p}{dt}$	$\frac{d\tilde{P}_j}{d\frac{P_p}{P_j}}$
17	1.2	577	1330	495	653	0.635	1.31	1140	1.510
18	2.0	419	1533	495	525**	0.000	0.00	1730	0.000
19	1.2	500	$\frac{635}{1465}$*	515	667	0.618	1.29	$\frac{450}{1890}$*	0.288
20	1.2	542	1547	528	709	0.655	1.34	1560	0.288
21	1.2	645	1624	486	655	0.656	1.34	1750	0.288
22	2.0	493	1388	510	510	0.000	0.00	1500	0.000
23	1.4	466	$\frac{669}{1336}$*	468	563	0.527	1.20	$\frac{615}{1670}$*	0.148

* The pyrogen pressure-time behavior was designed to have a definite increase in slope midway in the run.

** Following pyrogen burn-out, the JATO pressure stabilized at 525 psia.

NOMENCLATURE

P_{pi}	- Pyrogen Initial Chamber Pressure	(psia)
P_{pf}	- Pyrogen Final Chamber Pressure	(psia)
$P_{j\text{-ave}}$	- Reference JATO Normal Chamber Pressure	(psia)
$\tilde{P}_{j\text{-max}}$	- Maximum JATO Chamber Pressure (normalized)	
$D_{B\text{-max}}$	- Circle Diameter Equivalent to Blocked Area	(in)
$\frac{dP_p}{dt}$	- Pyrogen Pressure Rise Rate (mean)	(psi/sec)
$\frac{dP_j}{d\ P_p/P_j}$	- Slope of normalized overpressure vs. pressure ratio	

presented in Figure 14 have been normalized in order to clarify the analysis. Table III contains a summary of the test data. An additional test at ϵ^* = 1.2 was conducted in which the JATO igniter was not used and the pyrogen unit alone ignited the JATO. This run showed that the JATO would overpressure in a similar manner as presented in Figure 14 (i.e., no unusually high or low pressures, peaks, saddles, etc.). Note that runs 1-16 include preliminary hardware tests for this effort and tests which are associated with other aspects of Project SPRITE.

DISCUSSION OF RESULTS

The analysis of the results presented in Figure 14 led to the plots presented in Figures 15 and 16. Referring to Figure 14, it is noted that for ϵ^* = 1.2, three of the four test runs produced data which have the same initial slope of the normalized JATO pressure, $\tilde{P}_j$ ($\tilde{P}_j = \frac{P_j}{P_j} = \frac{\text{JATO pressure}}{\text{average JATO pressure (no pyrogen flow)}}$), and the ratio of pyrogen to motor pressure, P_p/P_j. Run 17 gave a slope which cannot be explained except by possible misalignment or misplacement of the pyrogen or some malfunction of the pyrogen or motor which may have partially blocked the JATO throat causing a higher overpressurization rate. However, since the other three runs at the same ϵ^* gave identical slopes (Figure 14), the slope of Run 17 is discounted as an anomoly.

The maximum JATO equilibrium pressures encountered in the runs have a maximum deviation of $\pm$ 3% at ϵ^* = 1.2. This is considered to be very acceptable. Run 20, which shows a temporary low $\tilde{P}_j$, can be

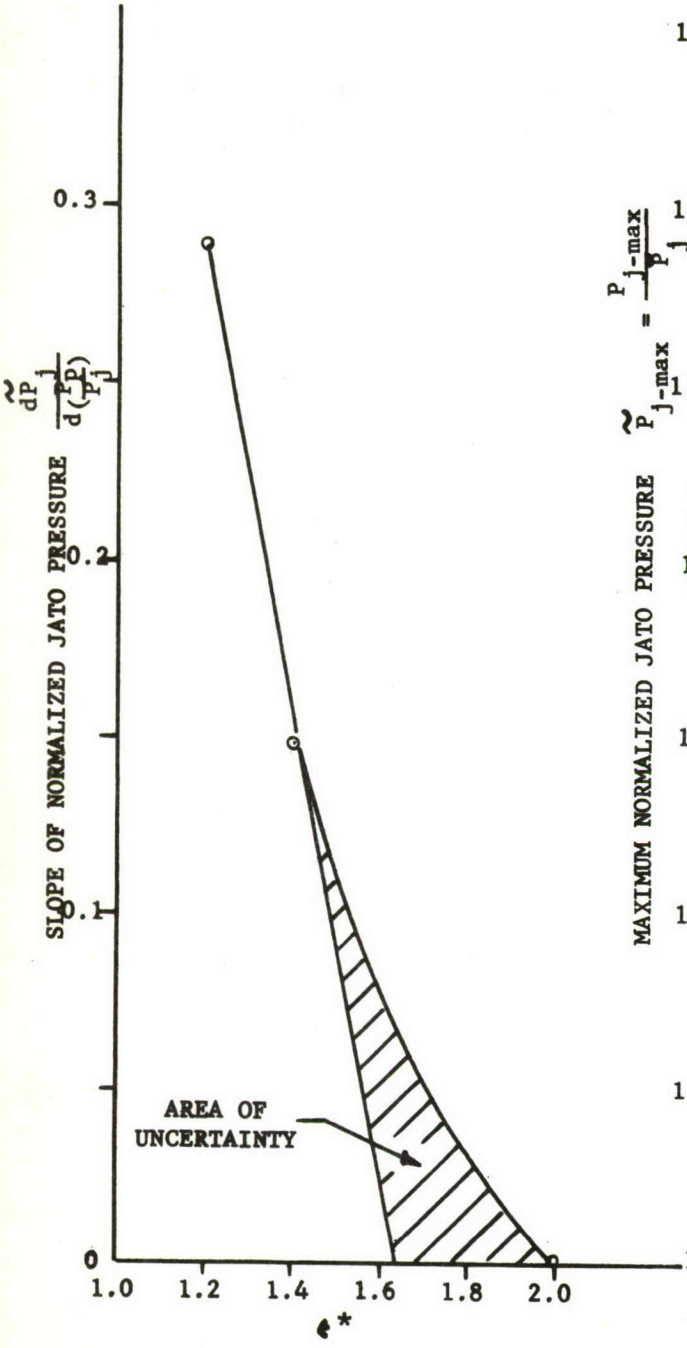


FIGURE 15 - RATE OF NORMALIZED JATO PRESSURE RISE WITH RESPECT TO THE PYROGEN TO JATO PRESSURE RATIO AS A FUNCTION OF PYROGEN LOCATION.

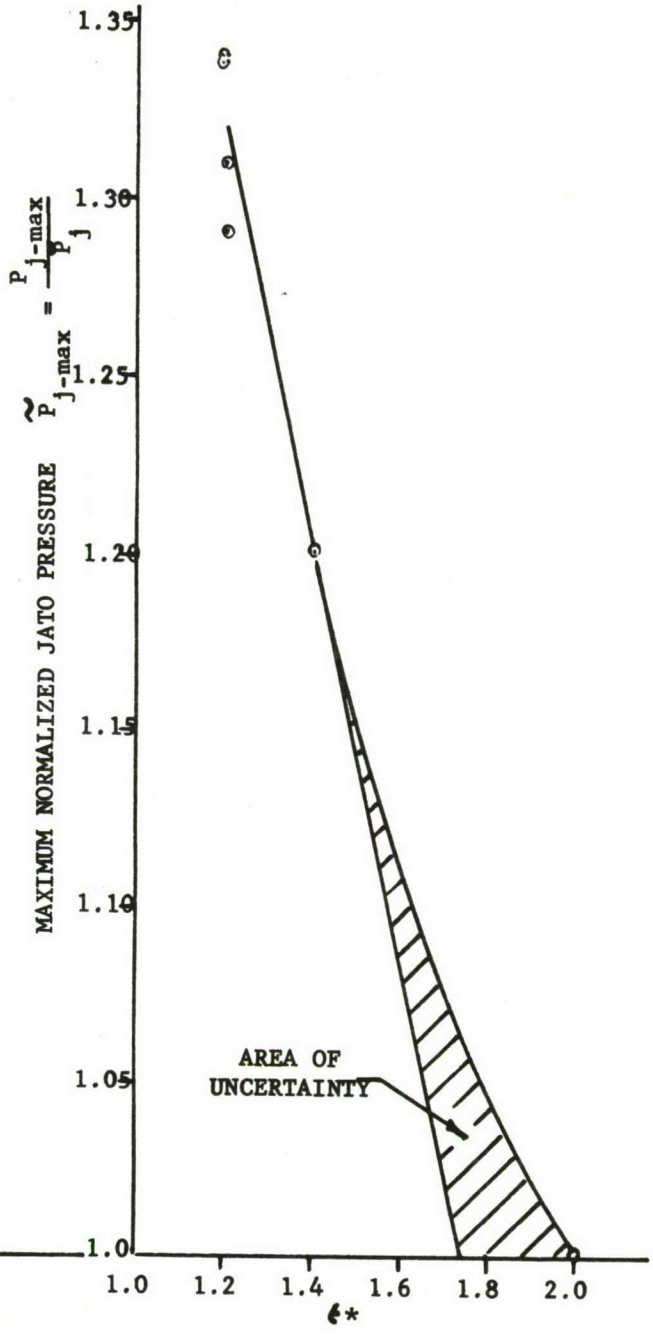


FIGURE 16 - RELATIONSHIP OF MAXIMUM OVERPRESSURE TO PYROGEN LOCATION.

explained by a lower and non-equilibrium JATO chamber pressure in the early portion of the pyrogen functioning time. This is very likely due to incomplete initial ignition. The pyrogen is presumed to have completed the ignition of the JATO causing the maximum overpressure, $\tilde{P}_{j\text{-max}}$, to rise to the same level as the other runs. Both the slope, $\frac{d\tilde{P}_j}{d(P_p/P_j)}$, and the maximum overpressure, $\tilde{P}_{j\text{-max}}$, are plotted against ϵ^* in Figures 15 and 16. Both figures show that the location of an igniter at an ϵ^* between 1.6 and 2.0 may prevent an overpressure and that a location at ϵ^* 2.0 will prevent any overpressure in this configuration. The area of uncertainty in Figures 15 and 16 is bounded by a linear variation and the known data point at $\epsilon^* = 2.0$. The expected overpressures at ϵ^* values between 1.4 and 2.0 should lie in this area. This finding is significant when compared to the UTC conclusions.

The stabilization of $\tilde{P}_{j\text{-max}}$ in each test is as yet unexplained since the theory indicates that as the pyrogen momentum increases the penetration should increase.

Table III lists values of $\frac{dP_p}{dt}$ during the period of pyrogen operation. It is interesting to note that a variation of $\frac{dP_p}{dt}$ does not affect the slope, $\frac{d\tilde{P}_j}{d(P_p/P_j)}$. This indicates that rapid pyrogen pressurization should not contribute to overpressurization of the motor being ignited.

In the introduction of Phase 2, the effect of propellant burning rate pressure exponent, n, on overpressurization was mentioned. Since the results of this program indicate the very real possibility of

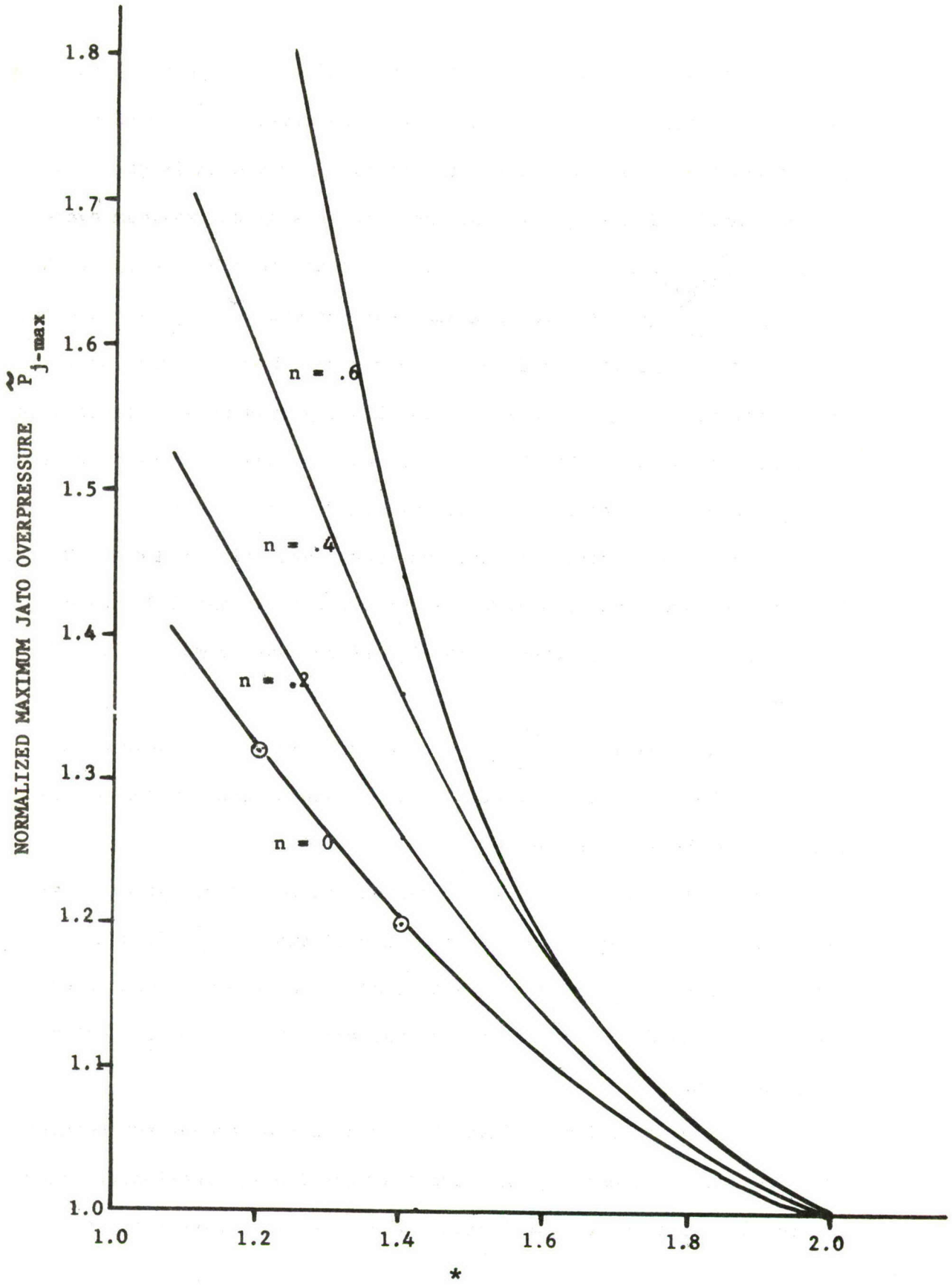


FIGURE 17 - EFFECT OF PROPELLANT BURN RATE EXPONENT, n, ON MAXIMUM OVERPRESSURE

overpressurization, Figure 17 has been constructed to show the effect of different n values on $\tilde{P}_j$ in this motor/pyrogen configuration.

CONCLUSIONS

The following conclusions were reached:

1. The possibility of main motor overpressurization due to the location, ℓ^*, of an aft end pyrogen igniter is very real.

2. For the test configuration used in this study, the location of an aft-end pyrogen at $\ell^* \geq 2.0$ will not cause an overpressure.

3. The rate of pressure rise in the pyrogen does not affect the state of overpressurization.

REFERENCES

1. Fullman, C. H. and Nielsen, F. B.,"(U) Theoretical and Experimental Investigations of Ignition Systems for Very Large Solid Propellant Motors," United Technology Center UTC 2012-FR2, Contract AF04(611)-7559, 8 May 1963.

2. Plumley, Arnold G., "Development of an Analytical Model to Determine Aft-End Igniter Design Parameters," Aerojet-General Corporation Technical Paper 136 SRP, Contract AF04(695)-350, March 1964.

LIST OF FIGURES

1 Schematic of Aft-End Pyrogen Ignition System

2 Theoretical Aft-End Pyrogen Penetration Correlation (After Plumley - Ref 2)

3 Penetration Study Apparatus

4 Experimental Configuration - Penetration Study

5 Data of Run 2

6 Data of Run 3

7 Data of Run 7

8 Sequence Photographs - Run 3

9 Data of Run 11

10 Throat Blockage Apparatus Assembly

11 Disassembled Throat Blockage Study Apparatus

12 Typical M-8 JATO Pressure-Time Trace Without Aft-End Pyrogen Firing

13 Typical Segment of JATO Pressure Time Trace, * = 1.2, With Pyrogen Firing; Pyrogen Pressure Time Trace

14 Normalized Test Data for Aft-End Pyrogen/M-8 JATO Tests

15 Rate of Normalized JATO Pressure Rise with Respect to the Pyrogen to JATO Pressure Ratio as a Function of Pyrogen Location

16 Relationship of Maximum Overpressure to Pyrogen Location

17 Effect of Propellant Burn Rate Exponent, n, on Maximum Overpressure

A RANGE SAFETY SYSTEM FOR MISSILES WITH TOXIC PROPELLANTS

Duane A. Haugen

Air Force Cambridge Research Laboratories (OAR)

L. G. Hanscom Field, Bedford, Massachusetts

ABSTRACT

A new and unique range safety problem was introduced at the missile test ranges by the advent of the TITAN II. The large quantities of toxic propellants that could become airborne in the case of a launch accident led range safety officials to rule in 1960 that TITAN II launches could be scheduled only with off-shore winds. At the request of AFBSD, we undertook a program to develop a means for continual quantitative specification of atmospheric capability to disperse toxic propellants in order to judge the advisability of launches with on-shore winds. We planned, conducted, and analyzed an extensive series of field diffusion experiments to develop the necessary quantitative relationships and concurrently had installed a computer-controlled meteorological data acquisition and processing system. The system has successfully supported every TITAN II launch at Cape Kennedy and Vandenberg AFB since March 1962. Its use has permitted demonstrably safe launches with on-shore winds. Emphasis in this paper is placed on a discussion of the diffusion prediction equations which were developed during this effort.

INTRODUCTION

The work to be described in this paper was carried out over about a two and one half year period starting early 1961. During this time, an intensive field program of turbulent diffusion experiments was accomplished at two major missile ranges, Cape Kennedy and Vandenberg AFB; a demonstrably reliable diffusion prediction equation was developed; and a new missile range instrumention system, the WIND System, was designed for and installed at each range. My purpose in this paper is to describe this effort, the reason for it, and to indicate in a general way the use that has been made of the results of the effort.

Before beginning, however, I would like to take this opportunity to state that a large number of people from several organizations contributed to the effort. Rather than attempt listing everyone who worked with us, I will simply emphasize that this was a team effort consisting of personnel from the Air Weather Service; The General Electric Company's Hanford Laboratories, Richland, Washington; Pan American World Airways, Inc. at Cape Kennedy; and The Martin Company at Vandenberg AFB. Key personnel of the Hanford Laboratories and of the Air Weather Service, particularly, worked with us in various phases of the effort, including publication of the experimental data and analyses of the data in a two volume AFCRL Research Report (5, 6). This report has not had unlimited distribution, but it is generally available to anyone who requests it. Therefore, I shall skip over many of the technical details which have already been reported, and concentrate instead on some aspects of the effort which were not discussed in the report.

THE TITAN - II RANGE SAFETY PROBLEM

The TITAN II poses a particular range safety problem that, at the time of its introduction to the missile test ranges, was unique and, in some cases, almost totally unfamiliar to range safety officials. The TITAN II is fueled by a mixture of hydrazine and UDMH (unsymmetrical dimethyl hydrazine) with nitrogen tetroxide as the oxidizer. Vapors evapprating from UDMH or boiling off from nitrogen tetroxide are toxic. The most serious problem is that of the oxides of nitrogen boiling off from pools of nitrogen tetroxide which has a boiling point of 70°F. In the completely fueled missile, there are about 14,000 gallons of UDMH-hydrazine and 17,000 gallons of nitrogen tetroxide. A propellant tank rupture during launch or fueling operations could thus expose large quantities of these fluids to the air. Estimates of the rate of boil-off or evaporation from a pool of nitrogen tetroxide range from about 45 to over 1000 pounds per minute depending on the surface area of the pool and the prevailing wind speed. Rates of evaporation from UDMH are considerably lower under identical conditions so that the limiting problem is to establish safe operating conditions for nitrogen tetroxide. At the time we were closely associated with the problem, the most probable rates of boil-off from nitrogen tetroxide were established at 500 to 1000 pounds per minute for Cape Kennedy, 250 to 500 pounds per minute for Vandenberg. The difference in probable rates arises because the launches at Kennedy are from gantries with containment of spilled fuels possible only in the relatively large area catch basins while silo-launches from Vandenberg permit the possibility of in-silo containment of spilled fuels.

Maximum permissible short term concentration levels that had been established for the toxic gases were 25ppm for nitrogen dioxide, the most prevalent oxide of nitrogen obtained, and 35ppm for UDMH. Given the expected rates of boil-off and ther permissible concentration levels, the problem was then reduced to describing the rate of atmospheric turbulent diffusion. In practical, realistic terms this problem had to be considered only for such times and meteorological conditions when the prevailing wind direction was towards inhabited areas close enough to the source to receive high concentrations of nitrogen dioxide. Let us therefore look briefly at the frequency of occurence of prevailing wind directions towards inhabited areas. These wind directions are defined as "unfavorable" for our purposes and are summarized by season in Table 1. These data were obtained by analysis of 28 months of continuous strip chart records of wind direction data for Cape Kennedy and 13 months of similar type record for Vandenberg AFB. Hourly mean values of direction were obtained from the charts for the analysis. It is evident from these data that only during the winter months does one have a reasonable expectancy of favorable wind directions. The other nine months of the year, one would expect fairly infrequent favorable wind directions, particularly at Vandenberg. As a matter of fact, unfavorable wind directions at Cape Kennedy during daylight hours only are as frequent as those at Vandenberg. Thus we see, in a general sense, that one must have some reliable means of quantitative determination of the toxicity hazard in order to establish TITAN II test schedules that are not dependent on the existence of favorable wind directions.

Table 1. Per cent frequency of occurence of unfavorable wind direction for TITAN II operations at Cape Kennedy and Vandenberg AFB. See text for description of data sources.

SEASON	CAPE KENNEDY	VANDENBERG AFB
Winter (Dec, Jan, Feb)	43	48
Spring (Mar, Apr, May)	64	87
Summer (June, Jul, Aug)	65	91
Fall (Sept, Oct, Nov)	69	69

When the TITAN II toxicity problem was first considered at the test ranges, concentration estimates were made using Sutton's equation for diffusion from a continuous point source (14). These equations are widely known by industrial hygienists, toxicologists, and meteorologists as being quite simple to use and very easily adapted to all conceivable configurations of pollutant sources. There are inherent difficulties, however, in using Sutton's equation; e.g.,

(a) there are four adjustable parameters in the equation which are related in a complex manner to meteorological parameters such as thermal stability and wind shear (3);

(b) experimentally determined values of these "Sutton" parameters have been found to vary considerably between sets of data obtained under similar meteorological conditions (4).

If one now couples with these difficulties the lack of appropriate meteorological data, wide ranges in concentration estimates might be expected. As an example, it can be shown that estimates of concentrations at distances of the order of one to six miles can differ by as much as one to four orders of magnitude (8).

In view of the above difficulties, it is not surprising that range safety officials and TITAN II personnel were unable to come to a mutually acceptable agreement about expected concentrations. As a result, it was decided that there would be no TITAN II operations at Cape Kennedy with onshore winds, the "unfavorable" wind direction as defined above. This practically eliminated the possibility of reliably scheduling daytime test launches at the Cape except for possibly the winter months.

It was at this juncture that AFCRL was requested to offer assistance by the TITAN II Weapons System Office. Our preliminary estimates of the scale of the problem indicated nominally three to five miles under daytime unstable conditions, ten to forty miles under nighttime stable conditions. But because of the uncertainties in our estimates, we proposed a program of experimentation and development designed to provide an operational solution acceptable to all concerned. Two very simple, but essential, ideas dictated our proposal: (a) the need to gather experimental data at the sites in order to develop operationally useful relationships between diffusion and meteorological parameters and (b) the need to develop an observational system for monitoring those meteorological parameters considered essential. Our goal was to improve from uncertainties of a factor of 10 to 10,000 to uncertainties appreciably less than an order of magnitude. Certainly such an improvement was called for if there was to be any meaning to the phrase: "operationally useful solution."

DISIGN OF DIFFUSION EXPERIMENTS

Our design of the field diffusion experiments was tempered by two fairly extensive sets of diffusion data already available to us in the literature; the Project Prairie Grass experiments, conducted in 1956 (1), and the Project Green Glow experiments conducted in 1959 (2). It was not known, however, whether the results of these experiments could be applied to the problem at hand because of: (a) differences in terrain features and vegetation cover; (b) the persistence of sea-breeze circulations at Cape Kennedy and Vandenberg; (c) the complicating factor of a well-developed, persistent subtropical marine inversion at Vandenberg. We therefore

designed experiments to highlight the significant differences we might find and to provide the minimal data necessary to develop relationships between diffusion and meteorological parameters.

In all the conceivable TITAN II accident cases except one, the spilled propellant will form a pool of liquid which will provide a continuous area source of toxic gas. The source height may be effectively raised above ground if the fuel and the oxidizer are spilled together or if the oxidizer is spilled with water. As a general rule, however, the ground level source presents the more serious diffusion problem and hence must be considered as the limiting case.

The one significant exception to the continuous area source accident is the case of a propellant tank rupture in which a large amount of propellant already in gaseous form will be immediately released into the atmosphere. In this case, the source type is termed an instantaneous volume source. There was not then, nor is there yet now, a satisfactory tracer technique for simulating an instantaneous source. It was necessary, therefore, to design field diffusion experiments for the continuous source case alone. In particular, we chose a ground source, emitting for a period of 30 minutes as the most serious conceivable accident situation that we could meaningfully simulate. The tracer technique used was the zinc sulfide particulate tracer technique developed by the Atmospheric Physics Group of the Hanford Laboratories. A detailed description of the technique may be found in a report on the Green Glow experiments (2).

For ease of reference, early in the effort the diffusion experimental programs were assigned nicknames to which we will refer here: At Cape Kennedy

the experimental program was nicknamed "Ocean Breeze"; at Vandenberg, "Dry Gulch".

The Ocean Breeze source point was located between TITAN - II Launch Pads 15 and 16 approximately 350 yds inland from the coastline. The diffusion course consisted of three arcs concentric on the source point at radii of 0.75, 1.5, and 3.0 miles. A schematic diagram of the Ocean Breeze sampling course is shown in Fig. 1. It should be noted that the course was purposefully oriented for experimentation during sea-breeze or on-shore wind circulations. The 1.5 mile arc was located to permit sampling just upwind of the industrial complex on the Cape during the sea-breeze situations. The 3.0 mile arc was located to permit sampling with north-northeast winds which occur fairly frequently during winter months. We could find no practical, economically justifiable way to sample over the Banana and Indian Rivers to the west of the Cape during sea-breeze conditions even though our preliminary estimates of the scale of the problem had indicated we might need diffusion data out to travel distances of five or six miles from the source.

The terrain at Kennedy consists generally of rolling sand dunes 10- to 20- feet high, widely covered by dense growths of palmetto and brushwood. The palmetto varies in height from about 2 to 5 feet; the brushwood from about 7 to 14 feet. Because of the height and density of the vegetation, the tracer samplers were placed at fifteen foot heights above local ground level, This represents a distinct departure from usual practices for field diffusion tests where samplers are placed at 1.5m above ground, defined as the average "inhalation level." Thirty samplers were installed at five foot heights to permit a check on the degree of vertical mixing within the cleared areas along Arcs 1 and 2. No significant difference was found between the fifteen and

five foot samplers as long as one considered exposures near the centerline or axis of the diffusing plume. We therefore proceeded to use the fifteen-foot concentration data for our analyses.

The terrain at Vandenberg is extremely complex. Significant terrain features consist of a broad mesa 200 to 300 feet in elevation sloping to the west with a 40 to 60 foot bluff at the coast line and rugged foothills and ridges inland. The mesa is cut up by several sharp, fairly deep ravines and has well-defined valley systems at the northern and southern edges. Both valley systems are oriented roughly along WNW to ESE lines; the northernmost valley (San Antonio Valley) bring fairly narrow with relatively steep sides, the southernmost valley (Lompoc or Santa Ynez Valley) being quite broad with gently sloping sides. Vegetation at Vandenberg consists of grass to a height of 1 to 2 feet, occasional clumps of brush 5 to 6 feet high and an occasional line of eucalyptus trees to heights of 50 to 80 feet.

The complexity of the terrain at Vandenberg led us to design two Dry Gulch diffusion courses; one up on the mesa, and one along the Lompoc Valley line. Both courses were oriented to sample the predominant on-shore sea-breeze circulation. The B-course, on the mesa, was located to permit sampling just upwind of the housing and industrial complex on the base. The source point was about 2600 yards inland from the coastline and located roughly halfway between TITAN II sites 395-C and 395-B. There were two sampling arcs on the B-course, at 1.43 and 3.52 mi.

The D-course was oriented along the Lompoc Valley line to obtain data upwind of private citizens living in the town of Lompoc, some ten miles

inland from the coastline. The source point was located about 1100 yards inland from the coastline in the immediate vicinity of TITAN II silo 395-D. There were three sampling arcs on the D-course at radii of 0.53, 0.93 and 2.93 miles.

Preparation and operation of the Ocean Breeze diffusion course was accomplished by personnel of Pan-American World Airways, Inc. The Dry Gulch courses were prepared and operated by personnel of The Martin Company, Vandenberg. Technical supervision of all field operations at both ranges was provided by personnel of the General Electric Company.

A minimal amount of meteorological data was obtained in conjunction with each diffusion experiment. Wind speed and direction were measured by a Belfort Company's Type M wind set located at each source point at a height of 12 feet above ground. Recording was a strip chart driven at a rate of 3 in/min for a nominal sixty minute period beginning with the tracer emission period.

Most of the temperature profile information was derived from three wiresonde measurements of temperature from the surface to 500 ft read at 50 ft intervals on each ascent and descent of the captive balloon. Some temperature difference data for Dry Gulch were also obtained from strip chart recordings of tower-mounted aspirated thermocouple measurements. The meteorological observations were obtained by Pan-American personnel at Cape Kennedy and by Air Weather Service personnel at Vandenberg. Reduction and analysis of all the meteorological data were accomplished at AFCRL.

There were 76 Ocean Breeze diffusion experiments conducted between 15 May 1961 and 26 March 1962. There were 109 Dry Gulch diffusion experiments

conducted between 12 June 1961 and 29 June 1962, 54 on the B-course, 52 on the D-course, 3 on the B-course and Arc D-3.

The diffusion experiments are summarized in Table 2 according to a subjectively-defined classification of thermal stabilities ranging from very unstable to very stable. Values of the temperature difference between 6 and 54 feet, denoted by ΔT, were used to classify each experiment.

The purpose of Table 2 is to emphasize the large number of experiments classified as moderately unstable. This is obviously a direct result of orienting the sampling courses to study turbulent diffusion with on-shore winds. While this strong bias in the data towards unstable conditions is of direct application to the TITAN II problem of operations during on-shore wind conditions, it is undesirable from the point of view of obtaining a general solution to the pollution problem for all stabilities. The method we adopted to overcome this difficulty was to develop a diffusion prediction equation utilizing the Prairie Grass data as well as the Ocean Breeze and Dry Gulch data.

The Prairie Grass experiments were conducted over flat prairie country near O'Neill, Nebraska in 1956. The vegetation was grass and hay which was kept mowed to a height of about two inches. The tracer material used was a gas, sulfur dioxide, emitted continuously for ten minutes and sampled along arcs at radii of 50,100, 200, 400, and 800m. A total of 68 experiments were conducted under a wide range of meteorological conditions. In particular, a wide range of thermal stabilities were sampled so that for these experiments, at least, no bias toward any particular meteorological condition exists.

Table 2. Ocean Breeze and Dry Gulch experiments summarized by stability classification. ΔT = temperature at 54 feet minus temperature at 6 feet.

	VERY UNSTABLE ($\Delta T \leq -3.0°F$)	MODERATELY UNSTABLE ($-3.0 < \Delta T \leq 0.0°F$)	MODERATELY STABLE ($0.0 < \Delta T \leq 3.0°$)	VERY STABLE ($\Delta T > 3.0°F$)	TOTAL NO. EXPERIMENTS
OCEAN BREEZE	2	58	16	0	76
DRY GULCH	12	89	8	0	109
TOTALS BY STABILITY CATEGORY	14	147	24	0	185

THE DIFFUSION PREDICTION EQUATION

A diffusion prediction equation is defined, for our purposes, as an equation which relates parameters of the diffusing cloud to meteorological parameters. In a more nearly correct sense, then, it is really an equation of specification permitting one to specify certain expected concentrations after being given a meteorological condition or situation. The distinction is made here merely to emphasize that our objective at the outset was to develop a "specification" equation, and not to predict what expected concentrations might occur in the future.

We approached the equation development work with one general qualification in mind. We wanted as a result an equation with a demonstrably reliable accuracy of sufficient degree to be of operational value and that was also as simple an equation as possible. The mathematical model we chose as a basis was the so-called Gaussian diffusion model (13) which for the centerline concentration takes the form,

$$\frac{C_p}{Q} = \frac{1}{\sigma_y \sigma_z \bar{U}} \qquad (1)$$

where C_p is the centerline concentration at some downwind distance from the source point; Q is the source strength; $\bar{U}$ is the mean wind speed; σ_y and σ_z are the standard deviations of the concentration distribution about the centerline in the y and z directions respectively at a given downwind distance from the source point.

Starting with a classical theoretical work for diffusion in a field of homogeneous turbulence by G. I. Taylor (15), one can show (7, 11) that to

a very good approximation

$$\sigma_y = f[\sigma(\theta), X] \tag{2}$$

where $\sigma(\theta)$ is the standard deviation of wind direction fluctuations as measured at the source point for example; X is the downwind distance from the source point.

Similarly,

$$\sigma_z = f[\sigma\varphi, X] \tag{3}$$

where $\sigma\varphi$ is the standard deviation of fluctuations in the wind elevation angle. The main difficulty one encounters in atmospheric turbulence near the ground, however, is that only $\sigma(\theta)$ may be reasonably assumed to be approximately non-variant in space; $\sigma\varphi$ varies markedly with height near the earth's surface (12). In particular, it generally increases with increasing height during unstable conditions so that one must actually measure $\sigma(\varphi)$ at several heights in the first few tens of feet of the atmosphere in order to try to obtain a representative value to use in Eq. (3). We therefore chose to use a "bulk" measurement of thermal stability for a descriptive parameter of diffusion rates in the vertical direction since diffusion is enhanced during thermally unstable conditions, inhibited during thermally stable conditions.

The diffusion parameter of prime concern was the so-called "normalized" centerline or peak concentration, C_p/Q, as defined in Eq. (1). Rather than trying to specify σ_y and σ_z and then use these to specify C_p/Q, we therefore attempted to develop a diffusion prediction equation of the following form:

$$C_p/Q = KX^a \overline{U}^b \sigma(\theta)^c \quad (\Delta T + k)^d \tag{4}$$

where ΔT, the "bulk" measurement of thermal stability, is the temperature difference between 6 and 54 feet (negative for unstable conditions); k is a positive constant added to ΔT to avoid raising a negative number to a power; K, a, b, c, and d are parameters of fit determined by least-squares regression techniques.

A number of different approaches were taken with the data in attempts to derive the most reliable equation for operational use. A detailed discussion of this work is contained in the AFCRL Report on Ocean Breeze and Dry Gulch (10). The approach found to be most satisfactory was to divide the data from Prairie Grass, Ocean Breeze, and Dry Gulch into two separate sets of data using ΔT as a ranking or ordering parameter in dividing the data. The two resulting sets of data were termed Ocean Gulch Grass for the dependent set of data, Dry Prairie Breeze for the independent set. A comparision of the observed values of the various parameters in the two data sets is shown in Table 3. It can be seen that the two data sets are indeed quite comparable; i.e. using ΔT as an ordering parameter did not bias the data sets.

The data sets further provide a reasonably large data sample for analysis by statistical techniques. As a matter of fact, to have a dependent and an independent set of turbulent diffusion data for 110 experiments in each set truly represents a "first-time luxurious experience for research meteorologists. In short, the opportunity to develop prediction equations on one set of data and then test the equations on another set of data has the very desirable value of being able to demonstrate the reliability of any particular equation.

It was found during the course of the analysis that standard statistical measures of the reliability of a particular prediction equation were unsatisfactory. In particular, the conventional computation of per cent reduction of error or variance between predicted and observed values of C_p/Q often gave negative or very low values. The reason for this is that linear regression techniques were used on the logarithm of Eq. (4) whereas standard errors were computed from actual predicted and observed values of C_p/Q. The non-linear transformation occasioned by taking logarithms thus negates the straightforward use of per cent reduction of error as a "yardstick" for the reliability of the prediction equation. It was therefore decided to let the percentage of cases that fell within as factor of two of the observed values (i.e., the predicted value is within 50-200% of the observed value) be the discriminator between various prediction equations. The choice of a factor of 2 is admittedly somewhat arbitrary and subjective but it does represent the best we think we can do with our present state of knowledge of the diffusion process. It is obviously a significant improvement over the uncertainties of one to several orders of magnitude that were discussed earlier.

The diffusion prediction equation obtained as the most reliable and accurate equation is:

$$C_p/Q = 0.00211\, X^{-1.96}\, \sigma(\theta)^{-0.506}\, (\Delta T + 10)^{4.33} \qquad (5)$$

where C_p is in units of gm/m^3; Q in gms/sec; X in m; $\sigma(\theta)$ in degrees; ΔT in °F. This equation, when tested on the independent set of data predicted 72% of the cases within a factor of 2 of the observed values, 97% within a factor of 4.

Table 3. Comparison of oberved values of indicated parameters in independent and dependent data sets.

PARAMETER	DEPENDENT SET	INDEPENDENT SET
C_p/Q (sec/m^3):		
Mean	0.000560	0.000496
Range	4.24×10^{-8} to 1.63×10^{-2}	8.14×10^{-8} to 1.67×10^{-2}
$\bar{U}$ (m/sec):		
Mean	4.40	4.16
Range	0.49 to 10.52	0.37 to 10.17
$\sigma(\theta)$ (deg):		
Mean	9.18	9.21
Range	2.62 to 31.70	3.60 to 28.52
ΔT (°F):		
Mean	-0.52	-0.52
Range	-3.78 to 10.55	-4.23 to 9.81
Number of Field Tests	110	110
Number of Arcs	321	321

It will be noticed that the mean wind speed is not included in Eq. (5). This is because it was found that inclusion of wind speed as a predictor did not improve the prediction accuracy over that obtained using only $\sigma(\theta)$ and ΔT as meteorological predictors. Hence, in the interest of simplicity, it is not included in the prediction equation. Eq. (5) has been programmed into the WIND SYSTEM installed at each range to provide the staff meteorologist and the range safety officer with continual values of expected peak concentrations in the event of an accident. Before discussing the WIND Systems, however, I would like to discuss the "demonstrated reliability" of Eq. (5).

Let us look first at Figure 3 in which a scatter diagram of predicted vs. observed values of C_p/Q are presented on a log-log plot. The dashed line in the middle represents perfect prediction; the first set of solid lines on each side of the dashed line is drawn for a factor of 2 between predicted and observed values; the second set of lines on each side, a factor of four. It can be seen from this graph that the computed "accuracy" or reliability of "72% of the cases within a factor of two" is maintained over five orders of magnitude of values of C_p/Q. That is, there is no apparent bias in the test of the equation on the independent sample towards any one of the individual series of diffusion experiments from which the sample was derived.

To emphasize this point, scatter diagrams similar to Fig. 3 are presented for the individual test series in Figs. 4, 5, and 6. We see that the same level of accuracy and lack of bias is found in all three plots. Considering the large difference in the terrain and vegetation at Kennedy,

Vandenberg, and O'Neill as well as the difference in experimental techniques between Prairie Grass and Ocean Breeze - Dry Gulch, we feel that Eq. (5) has been given the most stringent test possible and, hence, one can be confident of its stated reliability.

THE WIND SYSTEMS

The WIND System is a computer-controlled real-time data acquisition and processing system that was designed for and installed at the missile test ranges for direct application to toxicity problems at the ranges. A number of meteorological sensors to measure wind speed, wind direction, temperature, and vertical temperature differences were installed at selected locations on Cape Kennedy and Vandenberg for the purpose of providing a continual monitoring capability of the diffusion conditions. The sensor outputs are scanned, edited, and averaged over time intervals of 5, 15, 30 or 60 minutes. Primary system output components are a teletypewriter and tape punch. I would like to emphasize that the main function of the system is monitoring or observation. To predict such things as the value of the mean wind speed, $\sigma(\theta)$, or ΔT in the atmospheric boundary layer at some time in the future with any degree of confidence is a skill not yet attained within the meteorological profession. Therefore, the primary operational use of the system is to provide a tool for range safety officials and staff meteorologists for making a "go" or "no-go" decision with respect to TITAN II launch or fuel transfer activities and to have the information required to alter the decision when necessary. A secondary use of the system is to provide the basis for climatological summarizations of those

parameters pertinent to the low-level atmospheric diffusion problem. To frame these statements in other words, the solution of the prediction equation programmed into the computer which controls the system provides an answer to the question of how much pollutant may be expected at downwind distances in the event of an accident. The use of the equation within the framework of the WIND System provides answers to the questions of direction and speed of travel of the pollutant. The climatological summarizations of the WIND System data will provide quantitative knowledge of the ranges of the answers to these questions as dictated by the natural variability of the atmosphere.

It is perhaps worthwhile to mention that the climatological value of the WIND System at Cape Kennedy has already been realized by programs other than TITAN II. For example, NASA has expressed an interest in summaries of the data for application to potential toxicity problems in connection with some of the space boosters. The point to this is simply to comment that a development item carried out in direct response to and entirely supported by an Air Force weapons system has become a missile range tool of value and importance beyond the scope of the original problem.

The electronic engineering details describing the design, installation, and operation of the system will not be presented here since they have been published elsewhere (9,16). I will simply state that the WIND Systems have successfully supported every TITAN II launch at Kennedy and Vandenberg, including a few launches that provided pollution incidents.

SUMMARY

I have emphasized in this paper selected aspects of the Ocean Breeze - Dry Gulch programs. In particular, I have discussed at some length the reliability of a diffusion prediction equation developed for application to the TITAN II range safety problem. In concluding, I would like to take this opportunity to point out that this experience has very clearly defined for as areas where additional research is necessary to improve our capability to provide "operationally useful solutions" to future Air Force problems of this nature.

First of all, I do not wish to imply any sort of contentment with the stated accuracy of the diffusion prediction equation that I have presented. Surely, an uncertainty of 50 to 200% about an actual value cannot be accepted today as the very best we can ever hope to accomplish. I suspect, however, that several years of research will be required before this accuracy level is significantly improved, partly because obtaining experimental diffusion data is very costly, but primarily because the atmospheric turbulent diffusion problem is a very complex one far from being completely understood.

One of the most complex, least understood aspects of turbulent diffusion is the instantaneous source problem. It will be recalled that this type of source is also possible as a TITAN II launch accident. We were not able to provide any operational solution to this problem, however, because of a paucity of research effort, theoretical or experimental, on this problem. We are beginning at AFCRL this year to fund a program to develop a tracer technique suitable for studying the instantaneous source diffusion problem so that we can eventually provide an experimental technique and an

understanding of a physical process that will prove useful to TITAN II or some other weapons system or a space booster.

REFERENCES

(1) Barad, M. L., ed., "Project Prairie Grass, A Field Program in Diffusion", Vols I and II, Geophysics Research Paper No. 59, AFCRL, Bedford, Mass., 1958.

(2) Barad, M. L. and J. J. Fuquay, eds., "Green Glow Diffusion Studies," Vols. I and II, Geophysics Research Paper No. 73, AFCRL, Bedford, Mass., 1961.

(3) Barad, M. L. and D. A. Haugen, "A Preliminary Evaluation of Sutton's Hypothesis for Diffusion from a Continuous Point Source," J. Meteor., vol. 16, pp. 12-20, 1959.

(4) Haugen, D. A., Barad, M. L., and P. Antanaitis, "Values of Parameters Appearing in Sutton's Diffusion Models," J. Meteor., vol. 18, pp. 368-372, 1961.

(5) Haugen, D. A. and J. J. Fuquay, eds., "The Ocean Breeze and Dry Gulch Diffusion Programs, Vol, I", AFCRL Research Report, AFCRL, Bedford, Mass., 1963.

(6) Haugen, D. A. and J. H. Taylor, eds., "The Ocean Breeze and Dry Gulch Diffusion Programs, Vol II." AFCRL Research Report, AFCRL, Bedford, Mass., 1963.

(7) Hay, J. S., and F. Pasquill, "Diffusion from a Continuous Source in Relation to the Spectrum and Scale of Turbulence," Adv. in Geophysics, vol. 6, pp. 345-365, 1959.

(8) Holland, J. Z., "Qualitative and Quantitative Factors in Atmospheric Pollution Sampling", J. Franklin Institute, Monograph No. 4, 1958.

(9) Myers, R. F., "The WIND System (Weather Information Network Display)," in The Ocean Breeze and Dry Gulch Diffusion Programs, Vol. II, AFCRL Research Report, AFCRL, Bedford, Mass., pp. 23-47, 1963.

(10) Nou, J. V., "Development of the Diffusion Prediction Equation," in The Ocean Breeze and Dry Gulch Diffusion Programs, Vol. II, AFCRL Research Report, AFCRL, Bedford, Mass., pp. 1-20, 1963.

(11) Ogura, Y., "Diffusion from a Continuous Source in Relation to a Finite Observation Interval," Adv. in Geophysics, vol. 6, pp.149-160, 1959.

(12) Panofsky, H. A. and R. A. McCormick, "The Spectrum of Vertical Velocity Near the Surface," Quart. J. Roy. Meteor. Soc., vol. 86, pp. 495-503, 1960.

(13) Pasquill, F, Atmospheric Diffusion, D. Van Nostrand Co. Ltd; London, 1962.

(14) Sutton, O. G., Micrometeorology, McGraw-Hill, New York, 1953.

(15) Taylor, G. I., "Diffusion by Continuous Movements," Proc. London Math. Soc., Ser. 2, vol. 20, pp. 196-211, 1921.

(16) Taylor, J. H. "Operational Use of WIND Systems." in The Ocean Breeze and Dry Gulch Diffusion Programs, Vol. II, AFCRL Research Report, AFCRL, Bedford, Mass., pp. 49-73, 1963.

FIGURE TITLES

Figure 1. Ocean Breeze Diffusion Course, Cape Kennedy.

Figure 2. Dry Gulch Diffusion Courses, Vandenberg AFB.

Figure 3. Scatter Diagram of Observed vs. Predicted C_p/Q for Independent Data Test of Prediction Equation.

Figure 4. Prairie Grass Segment of Figure 3.

Figure 5. Ocean Breeze Segment of Figure 3.

Figure 6. Dry Gulch Segment of Figure 3.

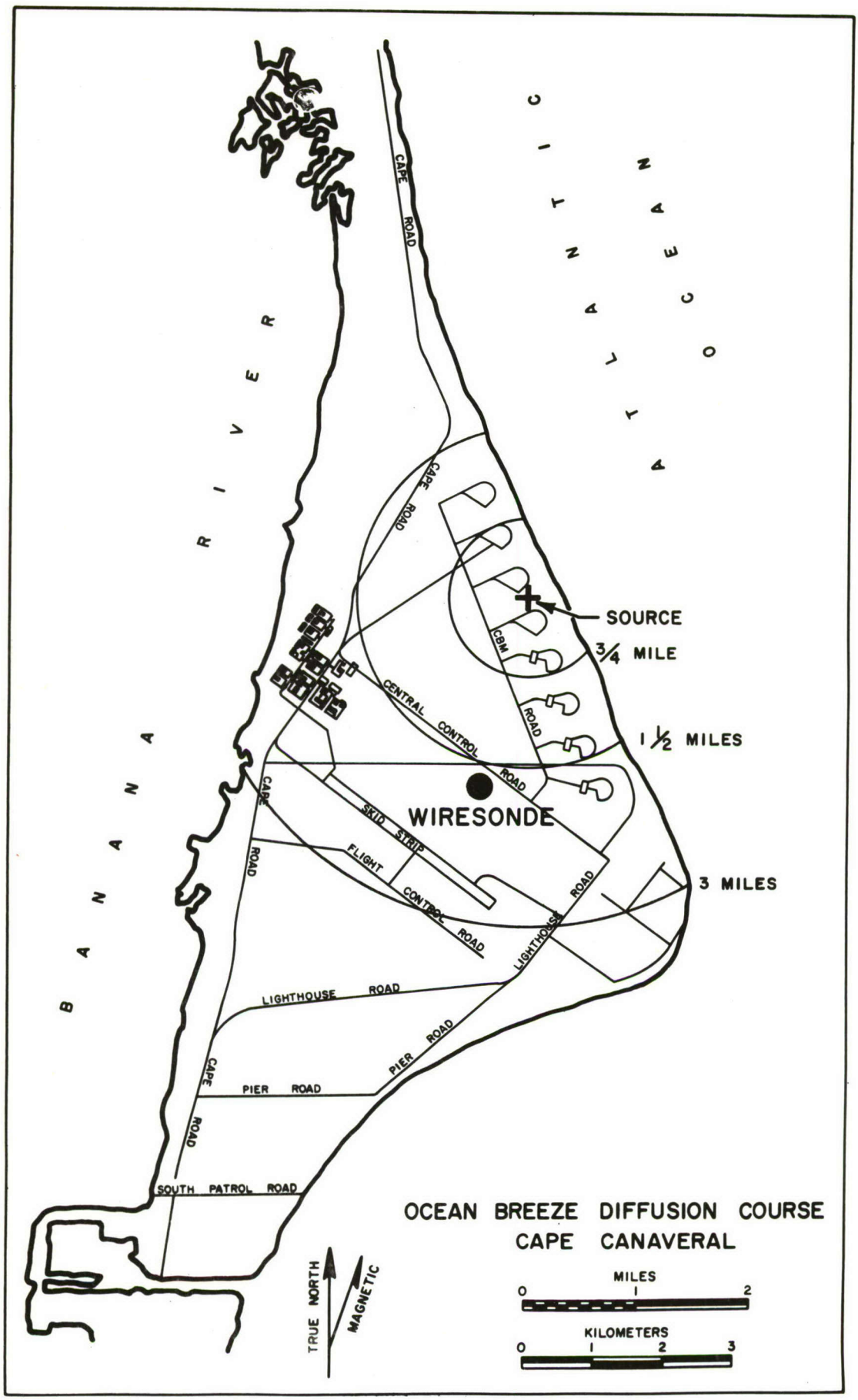

ATLANTIC OCEAN
BANANA RIVER
CAPE ROAD
CAPE ROAD
CBM ROAD
SOURCE
3/4 MILE
1 1/2 MILES
3 MILES
CENTRAL CONTROL ROAD
WIRESONDE
SKID STRIP
FLIGHT CONTROL ROAD
LIGHTHOUSE ROAD
LIGHTHOUSE ROAD
PIER ROAD
PIER ROAD
CAPE ROAD
SOUTH PATROL ROAD
TRUE NORTH
MAGNETIC
OCEAN BREEZE DIFFUSION COURSE
CAPE CANAVERAL
MILES
0 1 2
KILOMETERS
0 1 2 3

Fig. 1

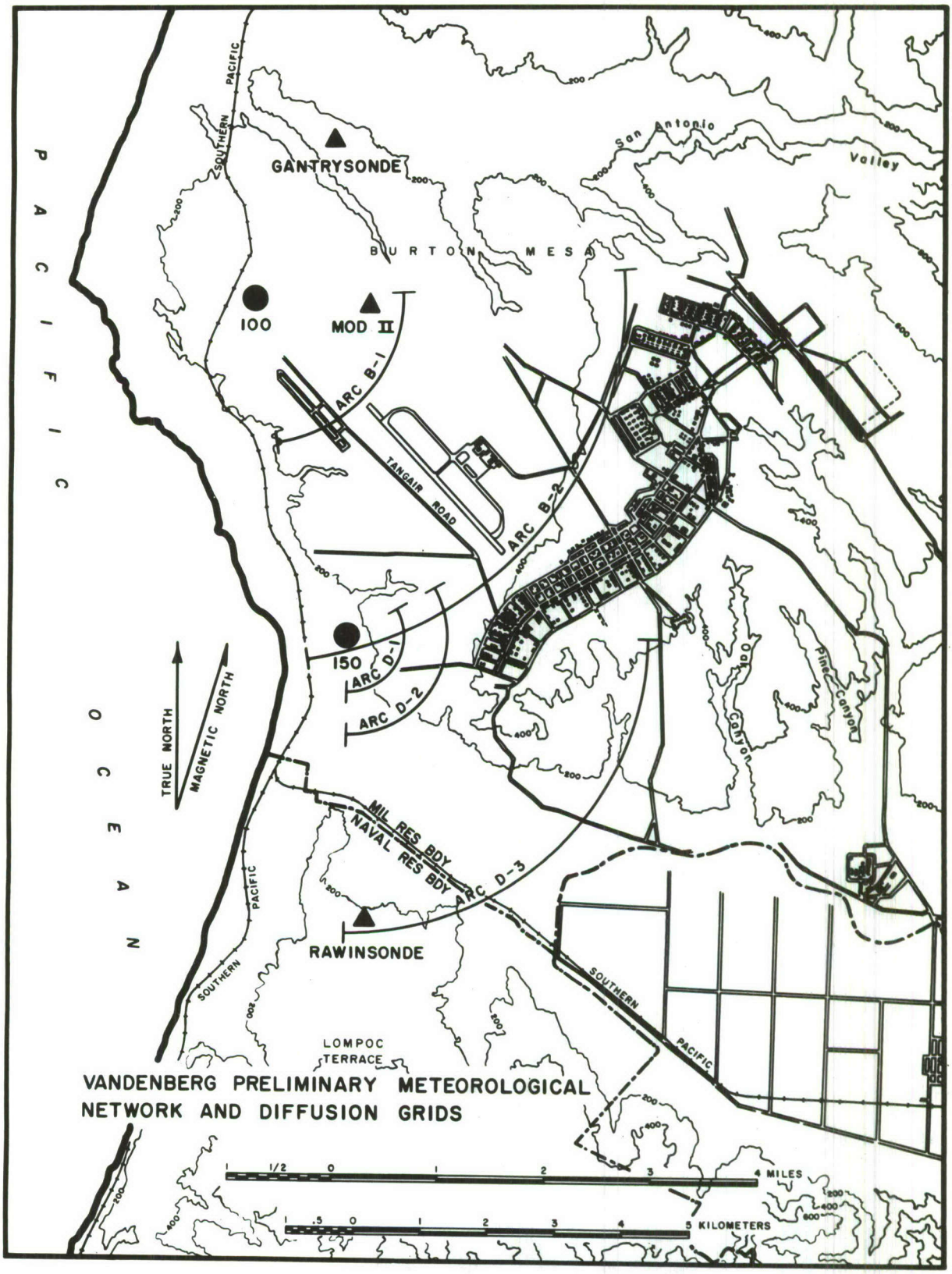

PACIFIC OCEAN
SOUTHERN PACIFIC
GANTRYSONDE
San Antonio Valley
BURTON MESA
100
MOD II
ARC B-1
TANGAIR ROAD
ARC B-2
150
ARC D-1
ARC D-2
TRUE NORTH
MAGNETIC NORTH
MIL RES BDY
NAVAL RES BDY
ARC D-3
RAWINSONDE
SOUTHERN PACIFIC
LOMPOC TERRACE
Pine Canyon
VANDENBERG PRELIMINARY METEOROLOGICAL NETWORK AND DIFFUSION GRIDS
1 1/2 0 1 2 3 4 MILES
1 .5 0 1 2 3 4 5 KILOMETERS

Fig. 2

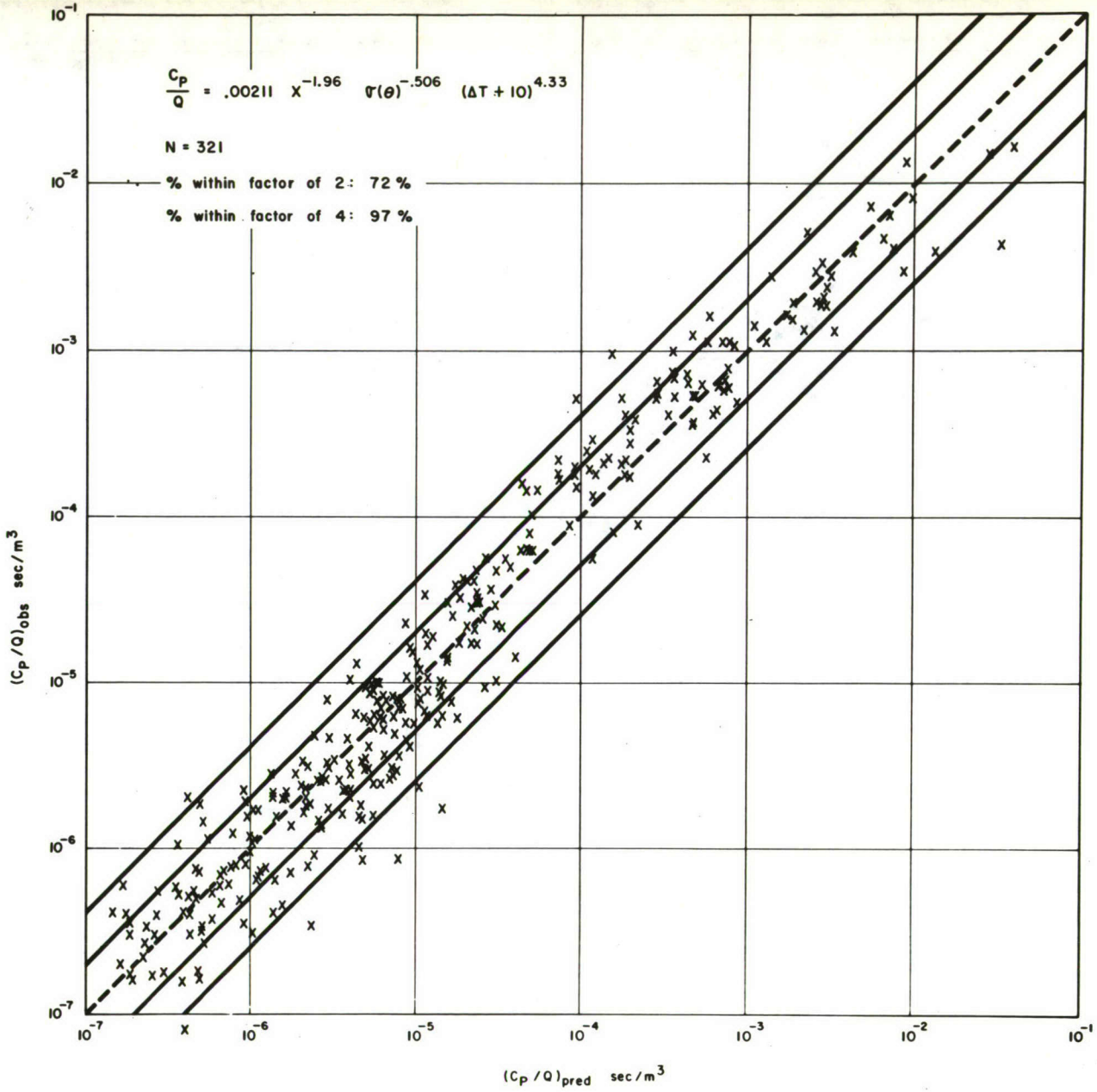

$\frac{C_p}{Q} = .00211\ X^{-1.96}\ \sigma(\theta)^{-.506}\ (\Delta T + 10)^{4.33}$
N = 321
% within factor of 2: 72 %
% within factor of 4: 97 %
$(C_p/Q)_{obs}$ sec/m^3
$(C_p/Q)_{pred}$ sec/m^3
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}

Fig. 3

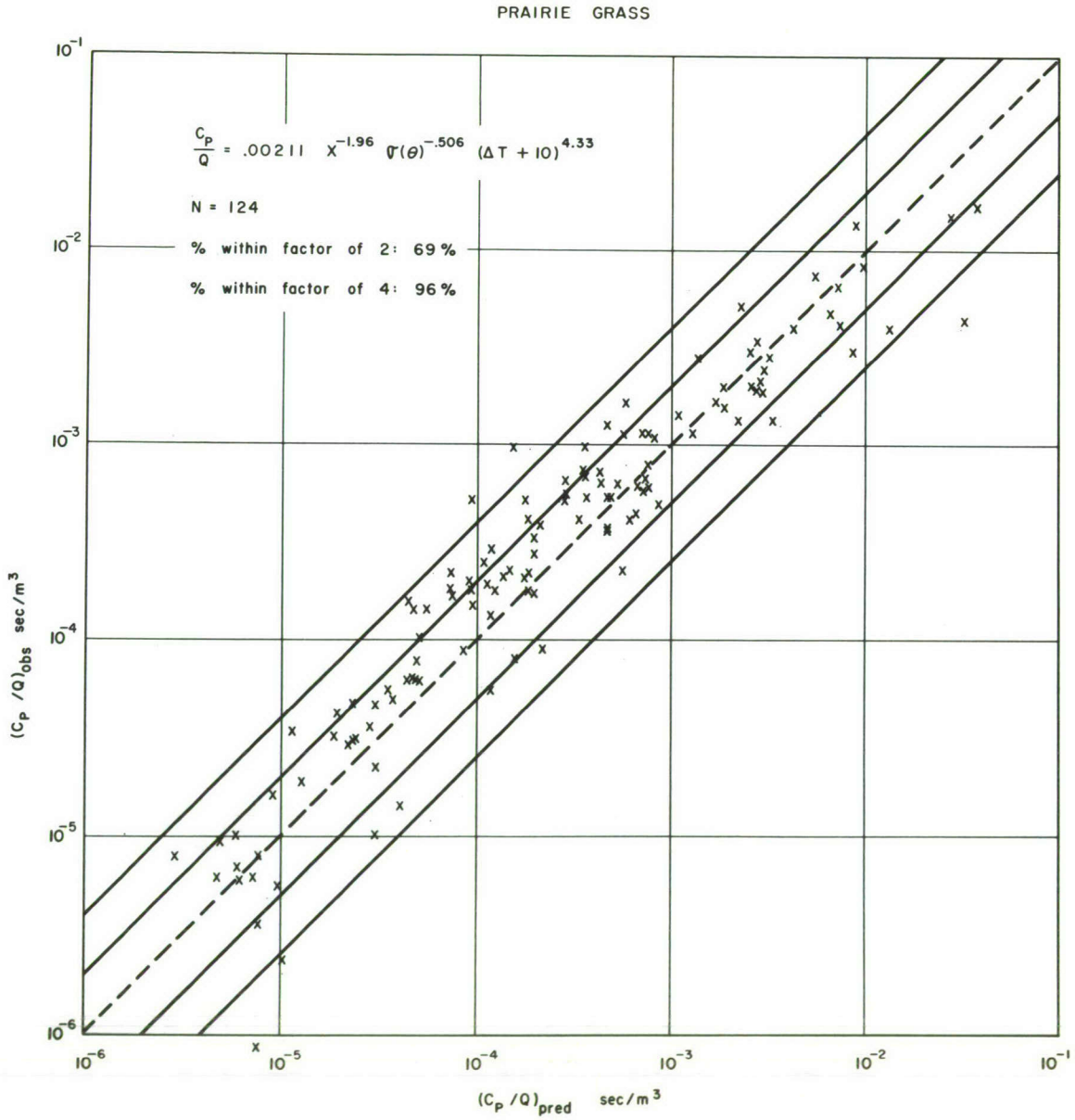

PRAIRIE GRASS
$\frac{C_p}{Q} = .00211\ x^{-1.96}\ \sigma(\theta)^{-.506}\ (\Delta T + 10)^{4.33}$
N = 124
% within factor of 2: 69 %
% within factor of 4: 96 %
$(C_p/Q)_{obs}$ sec/m^3
$(C_p/Q)_{pred}$ sec/m^3
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}

Fig. 4

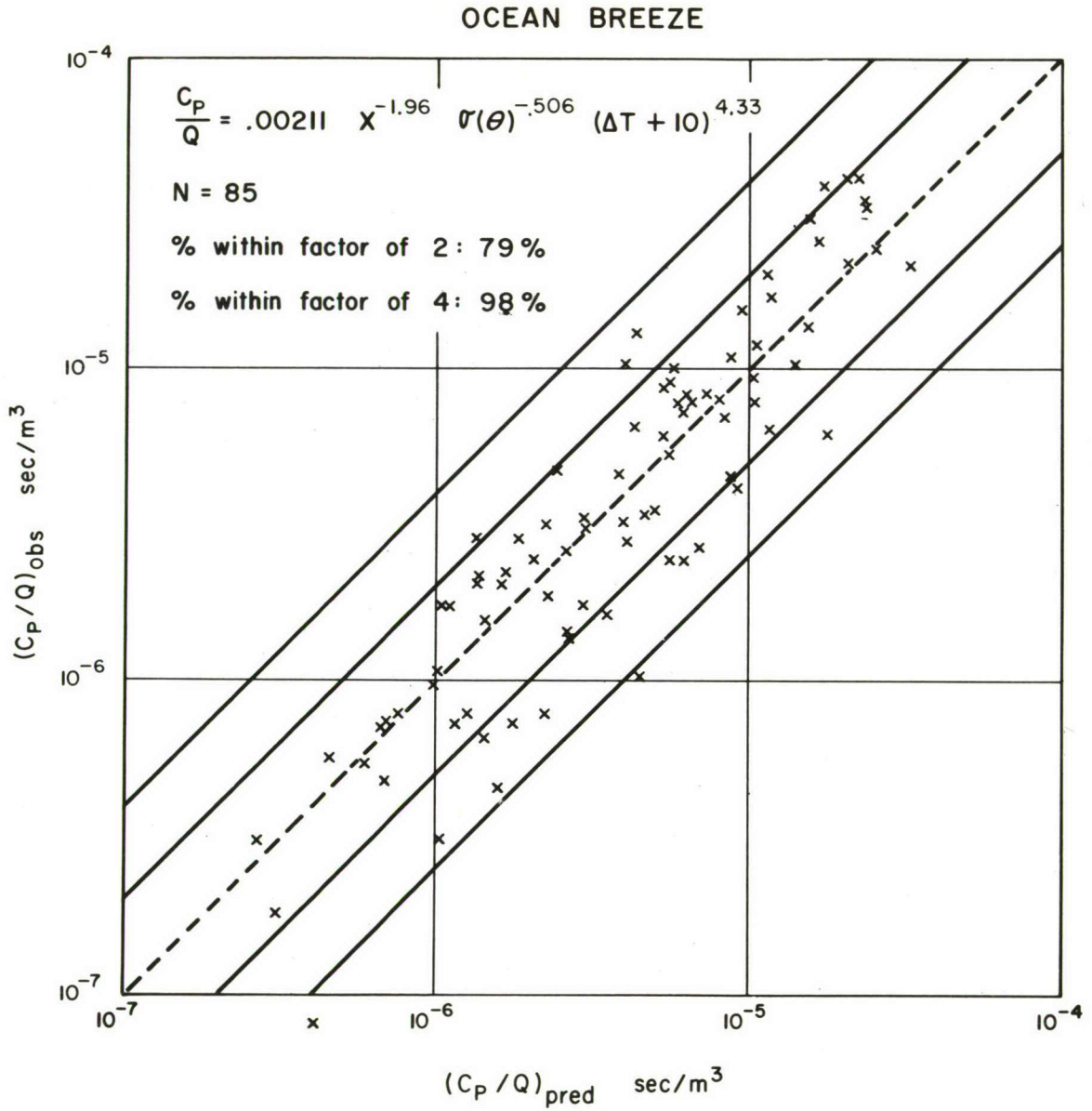

OCEAN BREEZE
$\frac{C_P}{Q} = .00211\ X^{-1.96}\ \sigma(\theta)^{-.506}\ (\Delta T + 10)^{4.33}$
N = 85
% within factor of 2: 79%
% within factor of 4: 98%
10^{-4}
10^{-5}
10^{-6}
10^{-7}
$(C_P/Q)_{obs}$ sec/m^3
$(C_P/Q)_{pred}$ sec/m^3

Fig. 5

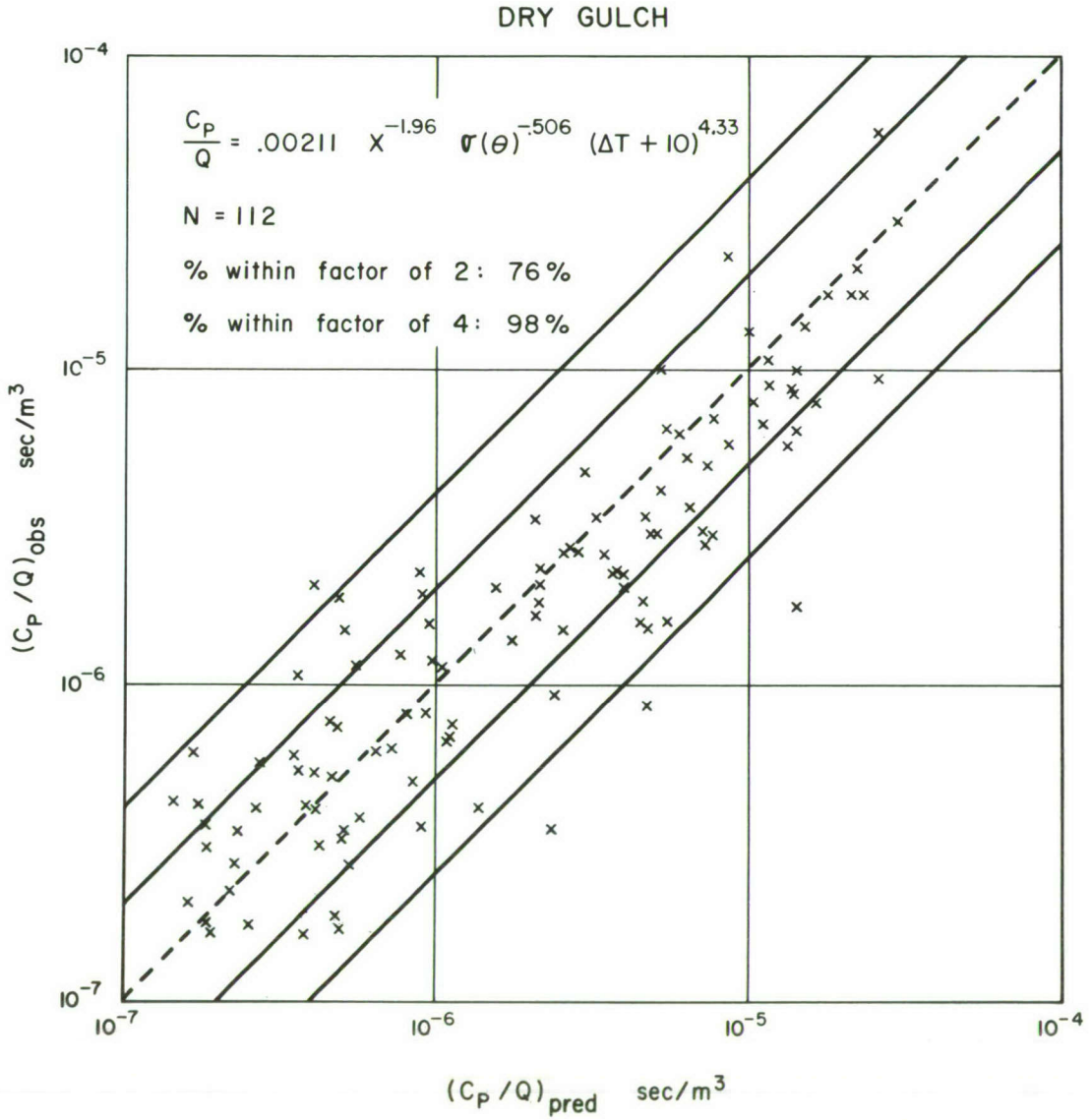

DRY GULCH
$\frac{C_P}{Q} = .00211\ X^{-1.96}\ \sigma(\theta)^{-.506}\ (\Delta T + 10)^{4.33}$
N = 112
% within factor of 2: 76%
% within factor of 4: 98%
10^{-4}
10^{-5}
10^{-6}
10^{-7}
$(C_P/Q)_{obs}$ sec/m^3
$(C_P/Q)_{pred}$ sec/m^3

Fig. 6

Mr. Emil J. Thomas, GS-13

Mr. Rodney Lee Clark, GS-11

Systems Engineering Group

Research and Technology Division

SEJDR

Wright-Patterson AFB, Ohio

(U) A NEW CONCEPT IN THRUST BEARINGS FOR OSCILLATORY MOTION

11 September 1964

UNCLASSIFIED

ABSTRACT

Title: A New Concept in Thrust Bearings for Oscillatory Motion

Classification: Unclassified

Co-Authors: E. J. Thomas, SEJDR

R. L. Clark, SEJDR

Category: Engineering Development

Conventional anti-friction bearings are primarily designed for full rotation applications, and the performance of these bearings in this type application is quite good. However, utilization of these bearings to accommodate oscillating motions has normally resulted in large reductions in bearing life. A particularly severe application of this type is the fully articulated helicopter rotor hub where oscillatory motion is provided about three axes to accommodate blade pitch change, flapping, and lead-lag requirements.

The development of the "laminated elastomeric thrust bearing" is described. The principle of elastic deformation is employed as opposed to anti-friction. Numerous laminae of thin stainless steel are bonded together with thin layers of elastomer. Under load, the elastomer is essentially under hydrostatic pressure. The steel laminae provide the strength to withstand the thrust load while the elastomer

provides for oscillatory motion by elastic deformation.

The development program includes preliminary design, construction, and life testing of prototype bearings for the purpose of investigating possible construction materials. Also included is the design and construction of a bearing adaptable to existing rotor hardware for use in a rotor whirl test and flight test program.

Future applications and required development effort are described. These applications require development of bearings designed to carry radial loading as well as thrust loading. An ultimate application would be a single elastomeric bearing designed to provide motion about three axes as required in fully articulated helicopter rotor hubs. The advantages to be realized are lighter weight, infinite life, elimination of lubrication requirements, and reduced maintenance requirements.

This development has been a combined in-house and contractor effort. Funds for this development were provided by the United States Army Transportation Research and Engineering Command.

(U) A NEW CONCEPT IN THRUST BEARINGS FOR OSCILLATORY MOTION

In the mid fifties, Air Force attention was directed to a new type bearing for use in applications where only oscillating motion is required. A patent was issued to Mr. W. L. Hinks for this bearing in 1958. (Ref. 1). Herein this bearing will be referred to as the "laminated pad" bearing.

Basically this bearing is constructed of numerous thin laminations of metal bonded together with thin layers of an elastomer. Under load, the elastomer is essentially under hydrostatic pressure. The metal laminations provide the strength to withstand the compressive load while the elastomer provides for oscillatory motion by elastic deformation. Several different geometric configurations were envisioned by the inventor; however, initial research and development effort was restricted to a simple flat washer arrangement of laminations. This type bearing has proven to be extremely well adapted to thrust bearing installations of the type found in helicopter rotor blade retention applications. Figure 1 illustrates this configuration and a typical helicopter installation. In Figure 1, the thickness of laminations is greatly enlarged for clarity. An actual bearing could have several hundred laminations, the number being determined by the required angle of oscillation. This new type bearing has the following advantages over

conventional anti-friction thrust bearings in oscillating motion applications:

(1) Less maintenance due to the fact that no lubrication is required.

(2) Longer bearing life.

(3) Lighter weight.

(4) Highly resistant to damage from dust contamination.

In order to clearly show the basic design principles, Figure 2 illustrates a simplified section of the laminated pad bearing consisting of two metal plates with a layer of elastomer between.

If a steady compressive load is applied to the illustrated bearing, the elastomer would deform as shown. If the thickness of the elastomer is reduced and an additional layer of metal added to make two laminations, the amount of extrusion, indicated by the dimension "x", will be reduced by approximately 50%. By holding the total amount of elastomer constant, while increasing the number of laminations, the extrusion of the elastomer can be limited to an extremely small amount. Therefore, by using very thin layers of elastomer and thin metal laminates, a dimensionally stable bearing can be manufactured.

Bearing geometry is determined in the following way. The cross section of the bearing is determined from the design axial load. Experience to date indicates that normal working stresses up to 13,000 psi are acceptable for a bearing

designed for a fatigue life in excess of 20 million cycles. The bearing can withstand momentary stresses from three to four times this magnitude without failure. The ID of the bearing is fixed by the requirement for transmittal of the design load by the supporting structural member that passes through the center of the bearing. With the bearing cross section fixed by the design compressive load, the required angular displacement capability is obtained by incorporating the required number of laminations. The angular displacement capability of each lamination is limited by the allowable shear stress in the elastomer at the OD of the bearing. For the early design work, the shear deformation angle (Figure 3) has been used to limit the shear stress to acceptable values. With elastomeric materials now available, shear deformation angles up to 27° have been found acceptable and working angles of approximately 13° will provide life in excess of 20 million cycles. At present, the distribution of stress between the elastomer and the metal in the area of the bearing OD is not clearly defined. The classical theory of elasticity can be used, but the actual stress in the elastomer can only be approximated. Finite deformations of the elastomer must be considered in a rigorous mathematical analysis of the "laminated pad" bearing. Such an analysis is currently being conducted under an Army contract by the Franklin Institute of Philadelphia. For the present, the design thickness for the

metal laminations is determined by calculating the hoop stresses produced in the metal laminations by the design compressive load. In addition, the optimum elastomer thickness for a given cross section bearing has been experimentally determined to be a function of the elastomer width (Figure 4). Based upon experimental data available, the ratio of metal width to elastomer thickness should be maintained at approximately 250.

Development of the "laminated pad" bearing began when Army interest was generated due to the potential helicopter applications. The U. S. Army Transportation Research and Development Command (TRECOM) provided funds for an Air Force monitored evaluation program.

The first phase of this program was an "in house" evaluation of four bearings produced by the Randolph Research Company which had been organized by the inventor. This "in house" evaluation was conducted at WPAFB and included fatigue and static testing. Figure 5 is a picture of the fatigue test machine used during this program. Two bearings are installed back-to-back in the machine. Axial load is applied by torquing down the nut visible in the photo. An aluminum washer, instrumented with strain gages, was used to measure the applied axial load. Oscillatory motion is applied through the lever arm. Bearing temperatures and oscillation speed were monitored. Several different

oscillatory angles and speeds were used during the fatigue test program.

This initial quantity of bearings was constructed of brass laminations bonded together with a natural rubber elastomer. The physical properties of these bearings were found to be good at room temperature; however, at temperatures below 0°F, the bearings stiffened rapidly to the point that they became completely unusable. Figure 6 shows a plot of the torque required to rotate the bearing vs. temperature. In light of the current military requirements for operation throughout a temperature range from -65°F to +160°F, the natural rubber bearings were not considered acceptable.

The next step in the development of the "laminated pad" bearing was to initiate a materials evaluation program with the primary objective of obtaining a combination of metal and elastomer that would be usable throughout a temperature range of -65°F to +160°F. The Franklin Institute was awarded an Air Force contract for this phase of development in mid 1961 and this evaluation was satisfactorily concluded in early 1964. Details of this program are reported in ASD Technical Documentary Report No. 63-769, "Development of the Laminated Elastomeric Bearing", Reference Nr. 2.

During the program, about 20 different elastomers were evaluated before a suitable elastomer was found. The final

materials combination selected was steel laminations bonded with an experimental silicone elastomer, General Electric XE-508. Figure 6 shows a plot of torque vs. temperature for this materials combination. Static and fatigue tests of test bearings produced excellent results. Fatigue testing was conducted in three machines of the type shown in Figure 7. One machine ran at room temperature, one at 160°F and the other at -65°F. Testing was conducted at oscillation frequencies of 300 cpm and 500 cpm. These frequencies were selected as being representative of actual helicopter applications.

The scope of the materials evaluation program did not permit an extensive study of construction methods; however, the success of the rather crude methods used indicates that precise control of the manufacture of the bearings can be maintained with a minimum of effort. Figure 8 illustrates the assembly methods used successfully in construction of the test bearings. The process used is referred to as the extrusion process. As shown, the elastomer is applied to the laminates with a small extrusion gun. Then a stack of laminates is placed in a die. Prior to curing, pressure is applied to force out excess elastomer. The extrusion of this excess elastomer results in a uniform bearing as shown in the photo in Figure 9. The specimen shown is a section from a typical test bearing magnified twenty five

(25) times. The specimen was prepared in the same manner as normal metallurgical specimens. The uniform spacing of the laminates indicates the success of this simple manufacturing process. The bearings were oven cured for several hours following the extrusion process.

Fatigue testing of the sample bearings has demonstrated that a fatigue life in excess of 20 million cycles is possible with proper design practice. Longer fatigue lives are believed possible with further materials refinements.

One of the significant characteristics of this type bearing is the mode of fatigue failure. The failure occurs in a very gradual manner with progressive breakdown of the bond between the elastomer and the steel. This failure progresses from the outer diameter to the center of the bearing. The failure is accompanied by an increase in torque required to oscillate the bearing. This failure is essentially "fail safe" as failure of the elastomer bond does not impair the load carrying capacity of the bearing.

Current effort is directed at procurment of "laminated pad" bearings adaptable for use in a modified UH-1B helicopter rotor. In this application, the bearing must carry a 90,000 lb compressive load while oscillating through a maximum oscillatory angle of $\pm 23^{\circ}$. The working angle for this application will be approximately $\pm 10^{\circ}$ during cruise flight.

In order to understand the requirement for this oscillatory angle change, the helicopter rotor must be examined.

In a helicopter, the primary flight control system consists of cyclic and collective pitch control of the rotor blades. The collective pitch control varies the angle of attack of each rotor blade uniformly. The cyclic pitch control system introduces a once per revolution sinusoidal pitch change into each blade so that the angle of attack of each blade at anytime is the algebraic sum of the collective and cyclic pitches. At each point around the azimuth of the rotor disk, the blade angle is varied according to the flight condition. In hover, all blades have essentially the same continuous angle of attack. In order to accelerate into forward flight, cyclic pitch is applied. In stabilized forward flight, the inflow to the rotor disk requires that cyclic pitch be applied so that the pitch of blades on the advancing side of the rotor disk is reduced while the pitch is increased on the retreating side. This variation in pitch is required to provide a net forward thrust while equalizing thrust on each side of the rotor disk. The ability to vary thrust across the rotor disk with the cyclic pitch control allows the pilot to control the attitude and direction of flight with the cyclic control system.

In order to accomodate the constant variation in pitch

of the blades, the helicopter rotor blade is usually attached to the rotor hub one of two ways. One method uses anti-friction bearings to carry the radial and thrust loads into the hub. The other uses a tension-torsion member which is constructed of either a continuous loop formed of wire or a bundle of steel straps. Both members are pinned at each end. The tension-torsion member carries the blade centrifugal load while the bending loads are carried by conventional roller bearings. The tension-torsion member must of necessity twist during pitch change. These tension-torsion members are fairly long in order to accomodate the required pitch changes. Figure 10 shows a size comparison of a "laminated pad" bearing and a wire tension-torsion member. Both are designed for use in a UH-1 helicopter rotor system.

The test program to be conducted using the "laminated pad" bearing equipped rotor will include a short flight test program along with a 110 hour whirl test of the rotor system. The whirl test is to be conducted on the RTD Rotor Whirl Test facility at WPAFB. This test program is scheduled for late 1964.

Future refinement and development of the "laminated pad" bearing is expected to result in numerous applications for this bearing. Advanced development is in progress which is expected to produce several geometric shapes suitable for many special applications. Some of these shapes are illustrated

in Figure 11.

Bearing "A" is designed to carry radial loading while allowing oscillatory motion about axis of the bearing. The laminates are arranged in the form of concentric cylinders. Some experimental specimens of this type have been tested with promising results.

Bearing "B" is a conical design with each lamination being in the form of a frustrum of a cone. Bearings of this type could be used in opposed pairs for applications where combined axial and radial loads must be transmitted. A pair of bearings of this type could be used in a helicopter rotor as the complete blade retention and pitch change bearing system.

Bearing "C" is a spherical design with each lamination being a portion of a spherical surface. This design is similar to a ball joint. Motion is permitted in any direction to a certain degree. One potential application is in a fully articulated rotor system to replace the present flapping and drag hinges in addition to the pitch change and blade retention bearings. Six or more conventional anti-friction bearings could be eliminated by such a spherical bearing. Only limited study has been directed toward this bearing to date; however, numerous potential applications exist for this bearing if it is developed.

Some of the potential applications for the "laminated

pad" bearing are in automobile suspension systems, in military and commercial tracked vehicles at the track connecting points, and in numerous other applications where oscillating motion is required. The advantages of low maintenance, longlife, and durability combined with the expected low cost of these bearings present a bright picture for the "laminated pad" bearing.

REFERENCES

1. U. S. Patent Nr. 2,900,182, "Static Load Bearings", 18 August 1959, Application 27 April 1955.

2. ASD Technical Documentary Report Nr. 63-769, "Development of the Laminated Elastomeric Bearing", 14 August 1963.

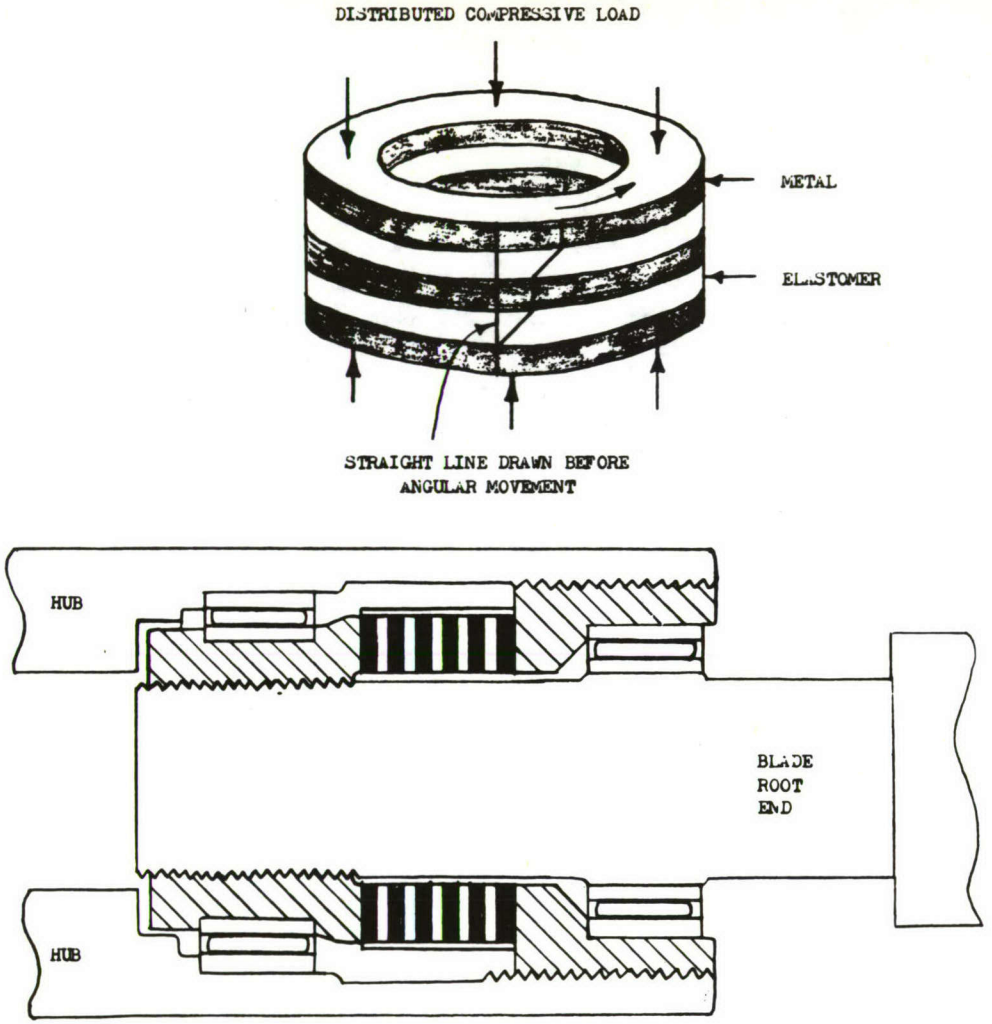


FIGURE 1

A Schematic of a "Laminated Pad" Bearing and a Typical Helicopter Application.

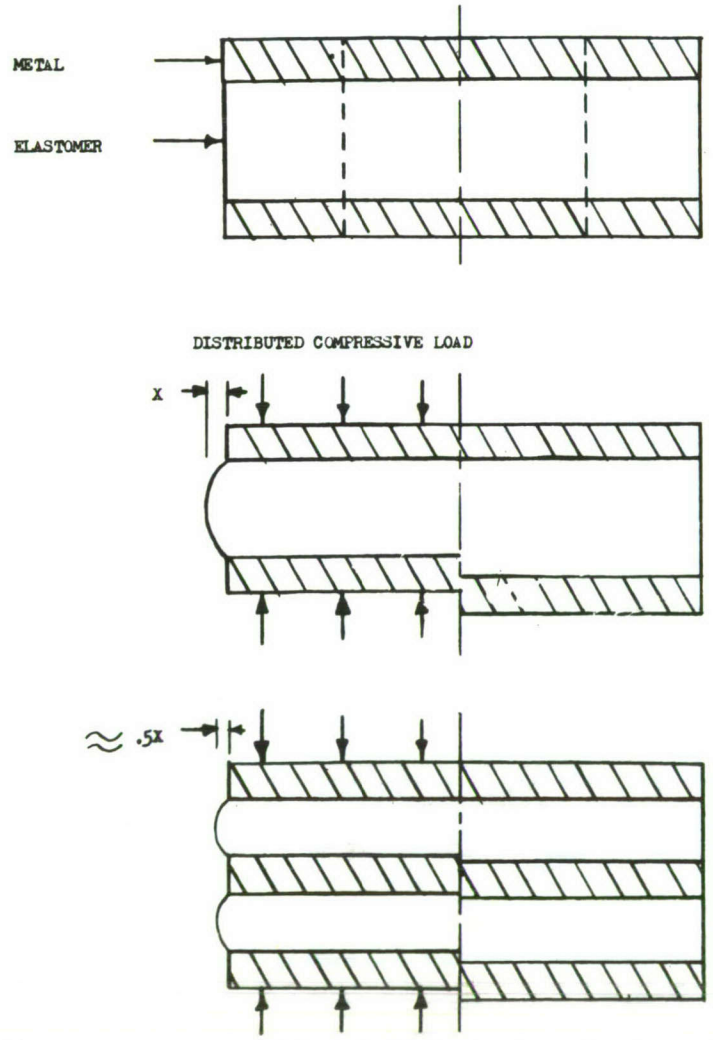


FIGURE 2

Elastomer Action Under Compressive Load

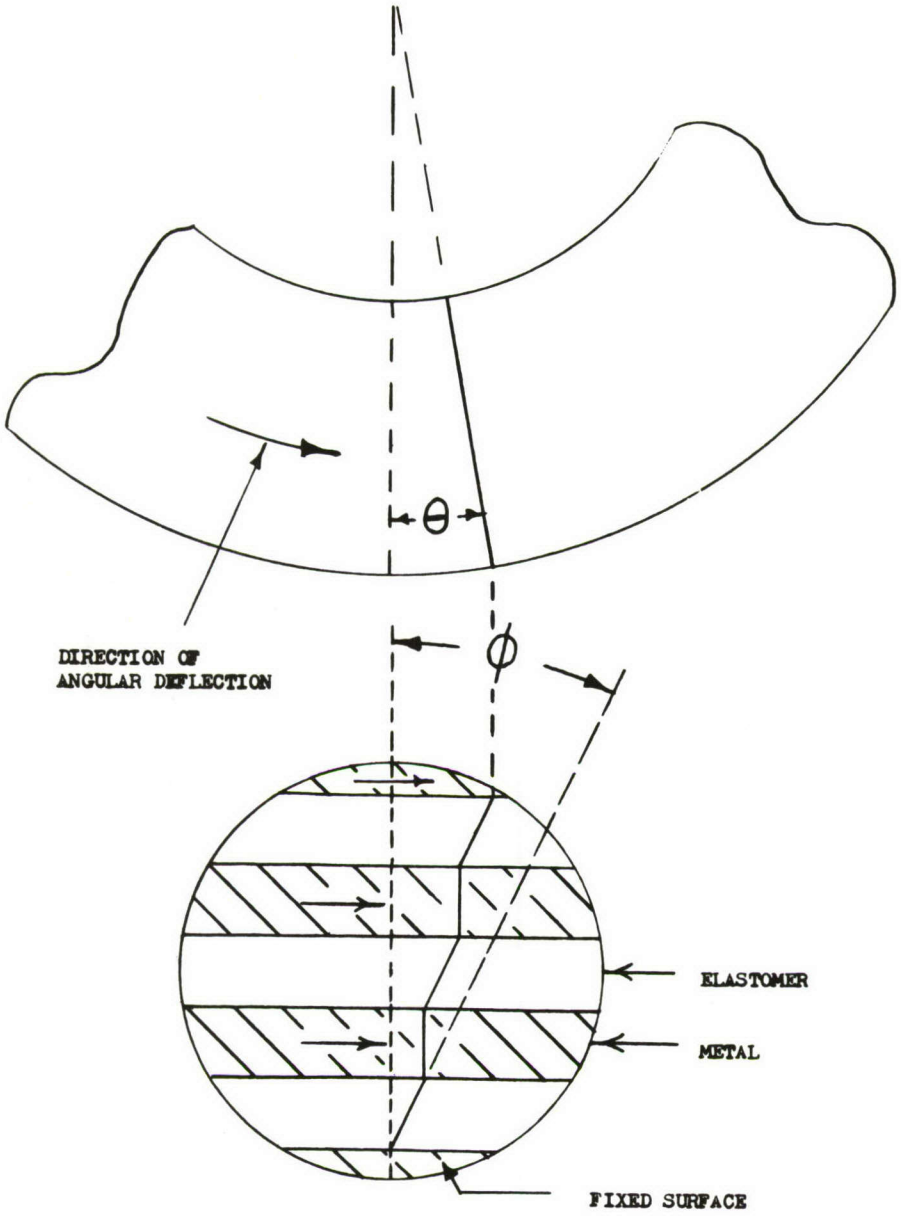
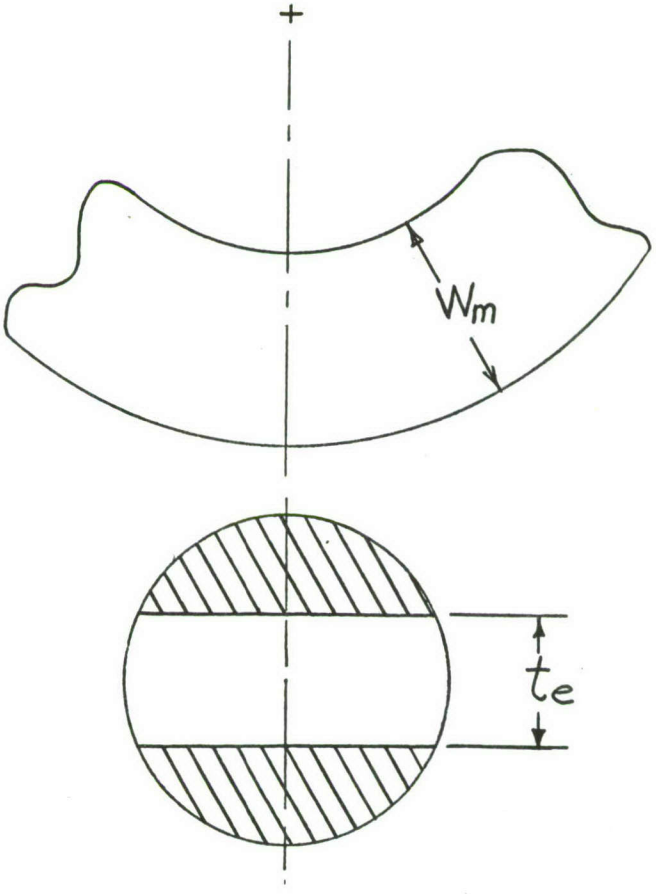


FIGURE 3

SHEAR DEFORMATION ANGLE (ϕ)

UNCLASSIFIED



RATIO OF $\frac{W_m}{t_e} \approx 250$

FIGURE 4

Elastomer Thickness

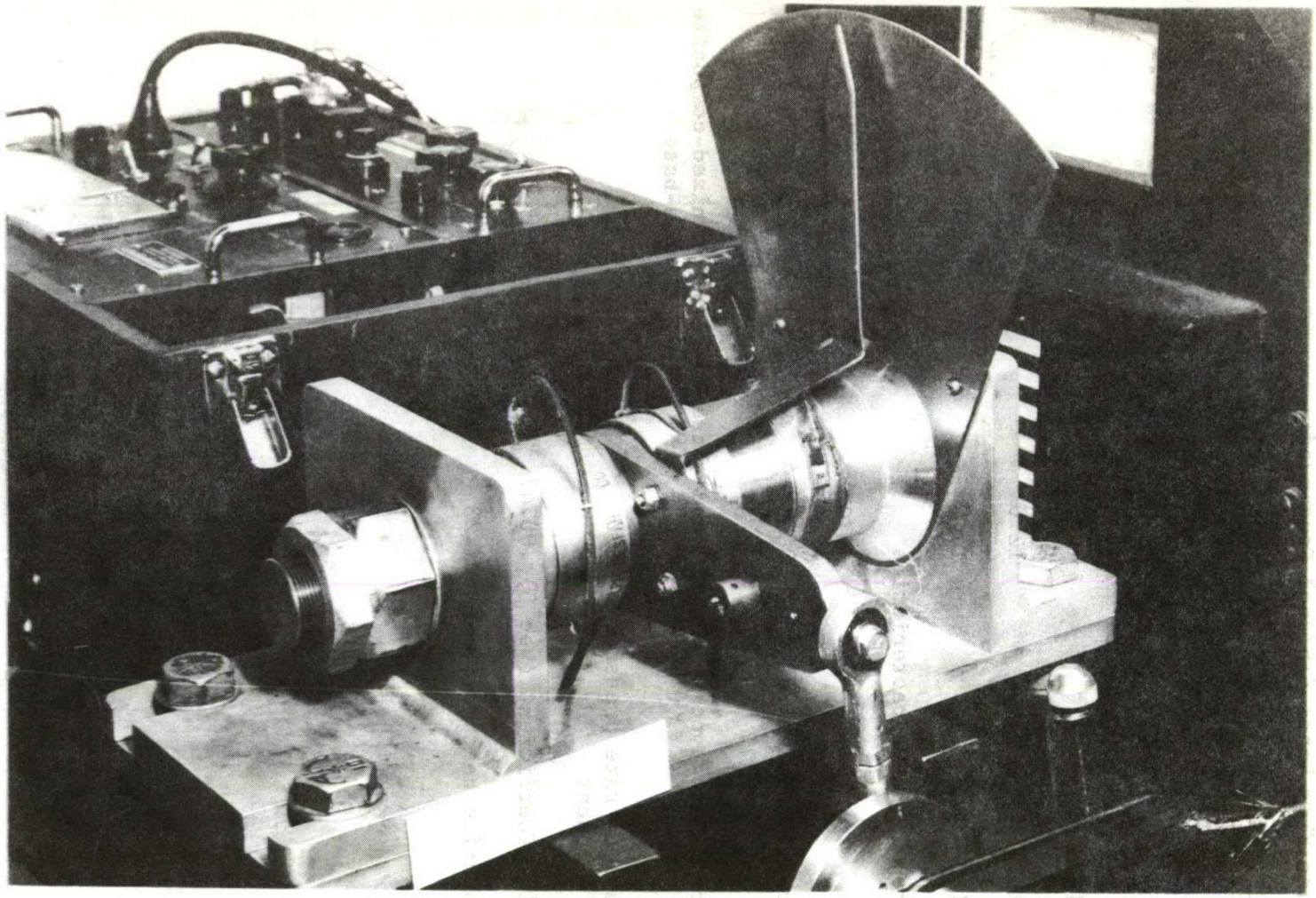


FIGURE 5

Original Fatigue Test Machine

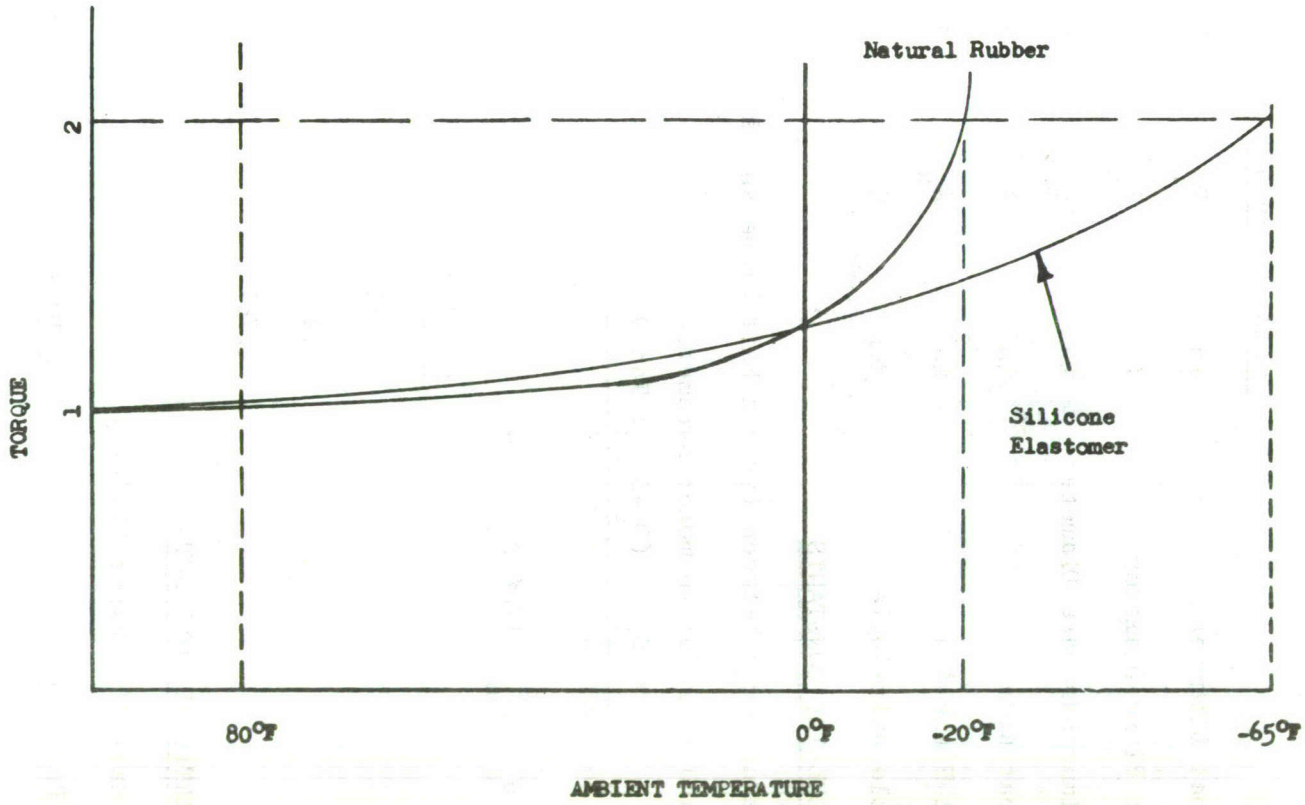


FIGURE 6
Torque vs. Temperature

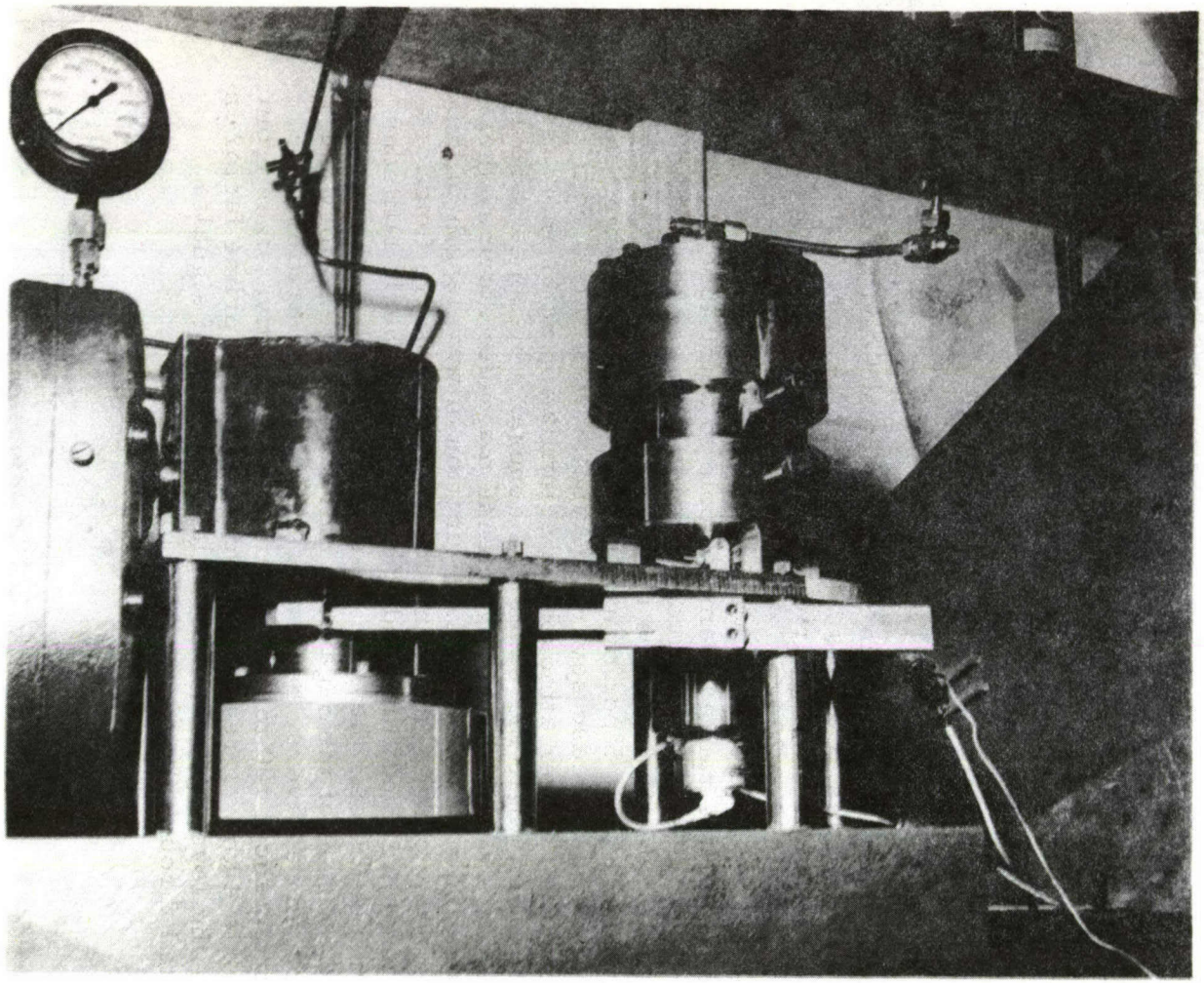


FIGURE 7

Bearing Fatigue Test Machine Used at Franklin Institute

UNCLASSIFIED

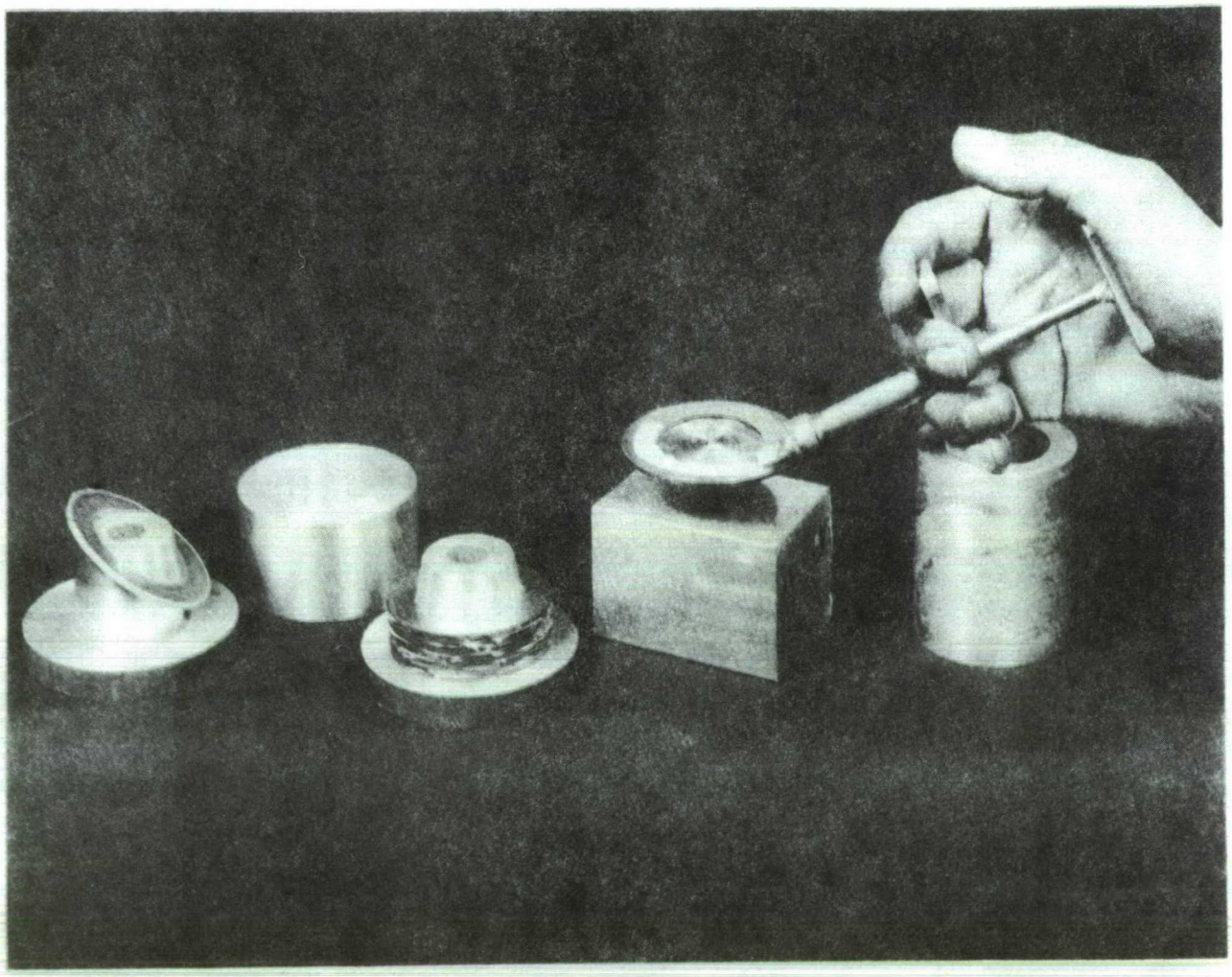


FIGURE 8

Experimental Assembly Method

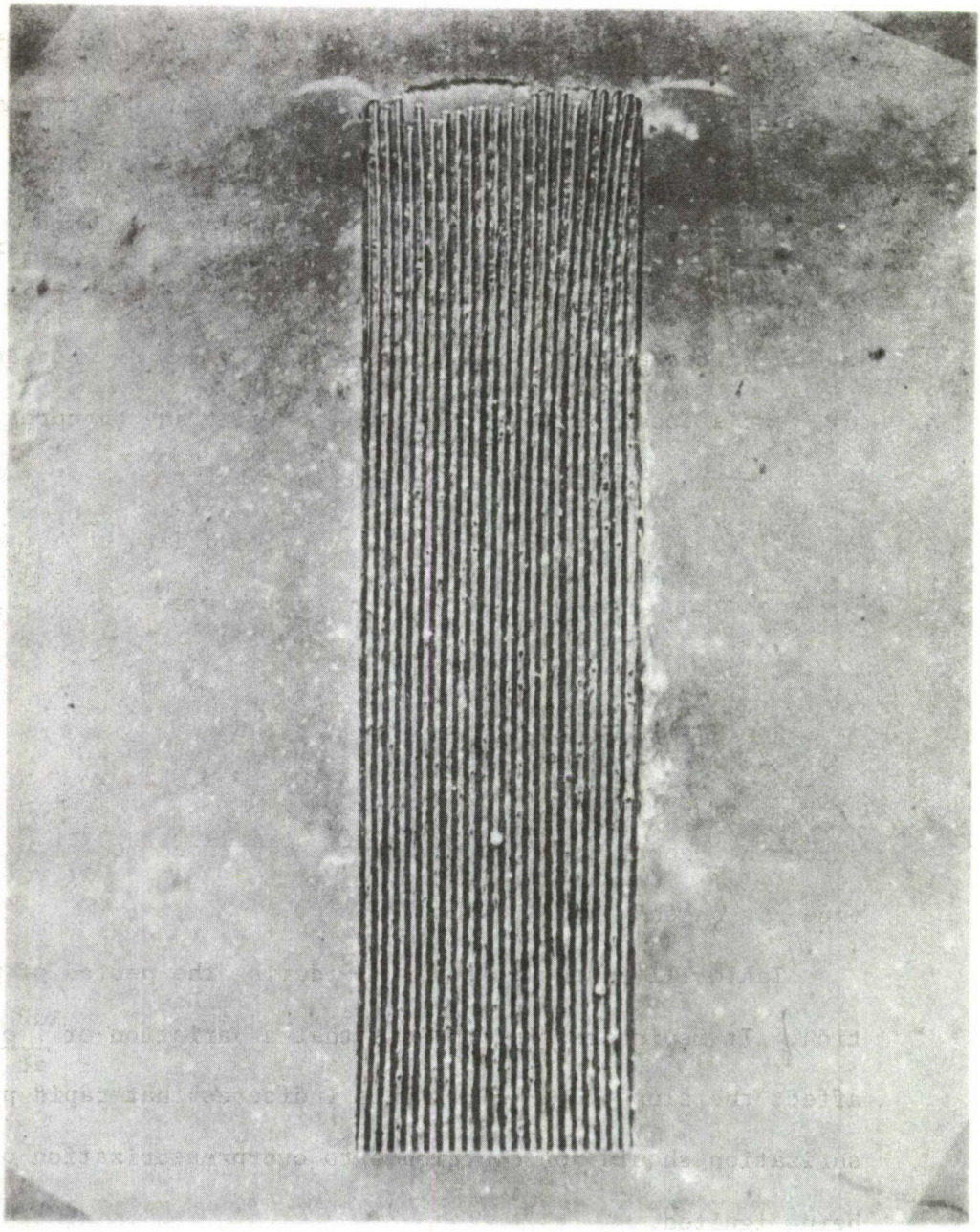


FIGURE 9

Magnified Cross Section of Bearing

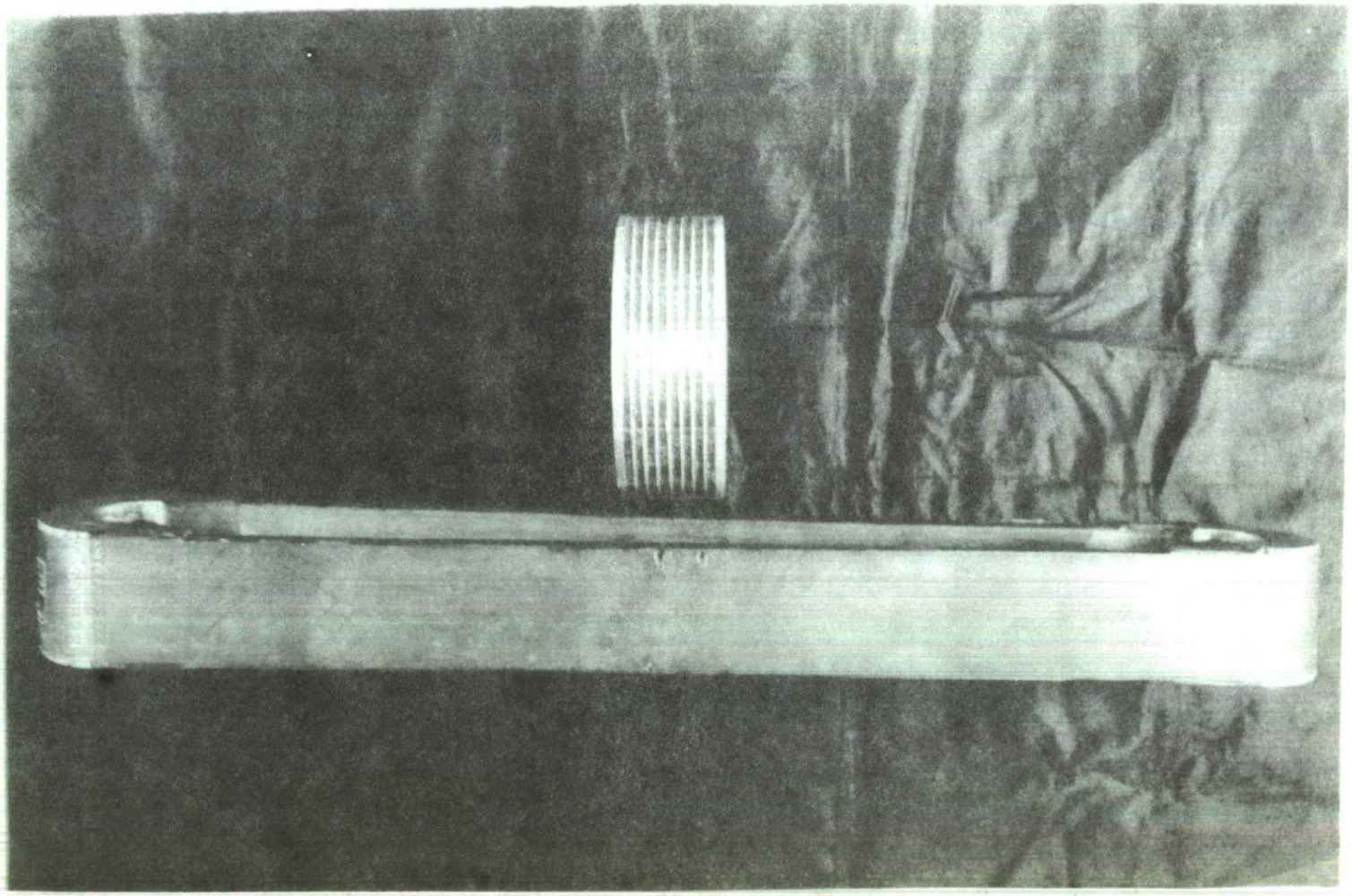
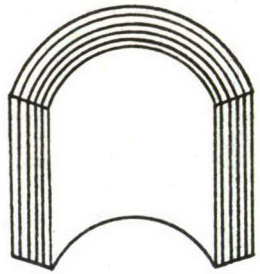
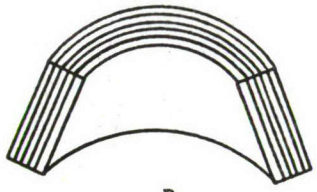


FIGURE 10

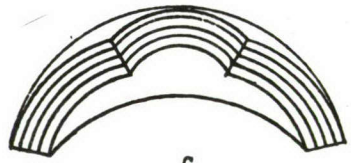
Size Comparison of "Laminated Pad" Bearing to Wire Tension-Torsion Member



A
RADIAL LOAD TYPE
BEARING



B
CONICAL BEARING



C
Spherical Bearing

FIGURE 11

Three Geometric Shapes for Future Bearings

SPACE USES OF THE EARTH'S MAGNETIC FIELD

Ralph B. Hoffman, 1st Lt, USAF
Thomas O. Haig, Lt Col, USAF

Space Systems Division
Air Force Systems Command

Title: Space Uses of the Earth's Magnetic Field

Security Classification: Unclassified

Authors: Ralph B. Hoffman, 1 st Lt, USAF, AO 3105659
Thomas O. Haig, Lt Col, USAF, 35757A

Category: Engineering Development

Abstract:

(U) SPACE USES OF THE EARTH'S MAGNETIC FIELD

For long life, oriented satellites, it is a virtual necessity to utilize passive attitude control systems. Fuel supplies for active control systems are excessive and result in untenable booster vehicle requirements. The more obvious passive systems which can operate in the near earth space environment are gravity gradient, solar sailing and magnetic torquing. The authors consider the third method, magnetic attitude control, in this paper. Where possible, the theory is illustrated with flight data. The paper is intended to provide a survey of: (1) Eddy Current Damping, (2) Hysteresis Damping, (3) Spin Control Using D.C. and Synchronous Motor Principles, (4) Two Degree Magnetic Attitude Control, and (5) Three Degree Magnetic Attitude Control (with inertia storage).

The paper includes heretofore unpublished data on the performance and design of a spin stabilized satellite with two degree magnetic attitude control. An extensive bibliography of articles on various "Space Uses of the Earth's Magnetic Field" is appended.

1.1 INTRODUCTION

A number of authors and satellite program offices have advocated or actually used the earth's magnetic field to achieve satellite attitude control, damping and referencing. This paper attempts to bring together many of these ideas and topics under a single, unifying title. The intent is to convey a comprehensive survey of "Space Uses of the Earth's Magnetic Field".

The sections on Eddy Current Damping, Hysteresis Damping, D.C. Motor Spin Control and Two Degree Attitude Control are illustrated with design and flight data. The section on a Three Degree Attitude Control System reflects current developmental thinking, though no flight hardward has been constructed. The section on a Synchronous Motor Satellite considers and then, for the present, dismisses the idea because of an unsolved damping problem.

The authors hope that the text will provide a ready reference for checking the practicality of proposed satellite systems using the earth's magnetic field. For example, several contractors have proposed an electromagnet for aligning a gravity gradient satellite with the vertical within a quarter orbit after launch. A closer look at the damping requirements, weight and power penalties for such a system, as developed in the text, suggests the scheme to be impractical.

The authors invite and encourage the reader to comment or criticize the paper for both the enlightenment of the authors and the improvement of the program for which they work.

2.1 THE EARTH'S MAGNETIC FIELD

The magnetic induction caused by a magnetic dipole moment is:

$$\bar{B} = -\bar{\nabla}\left(\frac{\mu_o}{4\pi}\frac{\bar{M}\circ\bar{R}}{R^3}\right) \qquad (2.1)$$

Where

$\bar{B}$ = flux density in Weber/Meter^2

μ_o = free space permeability ($4\pi \times 10^{-7}$ Henrys/Meter)

$\bar{R}$ = position vector to point of observation in Meters

$\bar{M}$ = dipole moment in Amperes-Meters^2

Ample evidence exists to justify the use of Eq. (2.1) to represent the earth's field between an altitude of one to ten earth radii. (1) (2) Below one earth radii, local surface anomalies and the offset of the earth's dipole from the geocenter causes appreciable deviations from the field predicted by a centered dipole model. Above ten earth radii, extra-terrestial phenomenon cause the dominate component of the magnetic field encountered. (3) At the present time, the extraterrestial phenomenon which give rise to the fields above ten earth radii are not understood well enough to permit long term prediction of the field. (Graph 2.1)

Using the earth centered dipole model, the magnetic field encountered in an arbitrary orbit will be derived. Fig (2.1) and (2.3) indicates diagrammatically the definition of the quantities listed in Table 2.1.

With this nomenclature and Eq. (2.1), the field in the xyz frame can be shown to be:

$$\bar{B}_{xyz} = \frac{\mu_o M}{4\pi R^3}\Big[\Big(-\tfrac{1}{2}\mathrm{Sin}\theta_m\mathrm{Cos}\psi_m+\tfrac{3}{2}\mathrm{Sin}\theta_m\mathrm{Cos}\psi_m\mathrm{Cos}2\varphi$$

$$+\tfrac{3}{2}(\mathrm{Cos}\theta\mathrm{Sin}\theta_m+\mathrm{Sin}\theta\mathrm{Cos}\theta_m)\mathrm{Sin}2\varphi\Big)\,\bar{i} \qquad (2.2)$$

$$+\Big(\tfrac{1}{2}\{\mathrm{Sin}\theta\mathrm{Cos}\theta_m+\mathrm{Cos}\theta\mathrm{Sin}\theta_m\mathrm{Sin}\psi_m\}-\tfrac{3}{2}\{\mathrm{Sin}\theta\mathrm{Cos}\theta_m+\mathrm{Cos}\theta\mathrm{Sin}\theta_m\mathrm{Sin}\psi_m\}\mathrm{Cos}2\varphi$$

$$+\tfrac{3}{2}\mathrm{Sin}\theta_m\mathrm{Cos}\psi_m\mathrm{Sin}2\varphi\Big)\bar{j}$$

$$+(\mathrm{Sin}\theta\mathrm{Sin}\theta_m\mathrm{Sin}\psi_m-\mathrm{Cos}\theta\mathrm{Cos}\theta_m)\bar{k}\Big]$$

Since θ and θ_m are known and essentially constant, Eq. (2.2) simplifies when numerical values are substituted. As an illustration, assume that a satellite mission requires a 500 NM, sun-synchronous, circular orbit. The period of the orbit is: (Ref. 4)

$$T = T_c(1+h/R_e)^{3/2} = 103 \text{ Minutes} \qquad (2.3)$$

Where T_c= Schuler Period = 84.347 Minutes

and the required inclination to maintain sun-synchronism is given by: (Ref 5)

$$\dot{\psi} = 0.01022\ (R_e/h)^2\mathrm{Cos}\theta \text{ rad/orbit period} \qquad (2.4)$$

or $\theta = 99.07^\circ$

Substituting these results into Eq. (2.2), the field for this particular orbit is:

$$\bar{B} = \frac{\mu_o M}{4\pi R^3}\Big[(-0.0997\ \mathrm{Cos}\psi_m+0.2991\ \mathrm{Cos}\psi_m\mathrm{Cos}2\varphi+\ 1.4045\ \mathrm{Sin}2\varphi)\bar{i}$$

$$+(0.4839 - 0.0157\ \mathrm{Sin}\psi_m- 1.4045\ \mathrm{Cos}2\varphi + 0.2991\ \mathrm{Cos}\psi_m\mathrm{Sin}2\varphi)\bar{j}$$

$$+(0.1544 + 0.1969\ \mathrm{Sin}\psi_m)\bar{k}\Big]$$

During the orbit period, T=103 minutes, the angle ψ_m will change by the amount $\Delta\psi_m=\psi_{m2}-\psi_{m1}=26^\circ$. The variation of ψ_m with time may be accounted for in one of two ways according to the accuracy required in the field description. If an exact description is not required, the mean value of ψ_m may be used with:

$$\psi_m = \frac{\psi_{m1}+\psi_{m2}}{2} \tag{2.6}$$

Where ψ_{m1} = ψ_m at time of nodal crossing

ψ_{m2} = ψ_m at time of next succeeding nodal crossing

The second method is to use the instantaneous value of $\psi_m = \frac{\pi}{12} t$ where t is in hours and ψ_m is in radians.

2.2 POLAR ALIGNED DIPOLE MODEL

If the earth's dipole moment is assumed to coincide with the earth's axis of rotation ($\theta_m=0$) and Eq. (2.2) simplifies considerably to:

$$\bar{B}_{xyz} = \frac{\mu_o M}{4\pi R^3}\left[\frac{3}{2}\mathrm{Sin}\theta\mathrm{Sin}2\varphi\bar{i} + \tfrac{1}{2}\mathrm{Sin}\theta(1-3\mathrm{Cos}2\varphi)\bar{j} - \mathrm{Cos}\theta\bar{k}\right] \tag{2.7}$$

2.3 TWO DIPOLE MODEL FOR A CIRCULAR ORBIT

For a circular orbit, an alternate derivation of the magnetic field may be made which simplifies subsequent discussion. The earth's dipole moment may be resolved into two components, one perpendicular to the orbit plane and the other lying in the plane. The total field is then the super-position of the fields caused by the two dipoles. From elementary physics, the planar field due to a dipole M_{xy} is

$$\bar{B}_N = \frac{\mu_o M_{xy}}{4\pi R^3} \quad (2\mathrm{Sin}\varphi)\bar{a}_n \tag{2.8}$$

$$\bar{B}_T = \frac{\mu_o M_{xy}}{4\pi R^3} \quad (\mathrm{Cos}\varphi)\bar{a}_T \tag{2.9}$$

The magnitude of the planar field is:

$$B_{xy} = \sqrt{B_N^2 + B_T^2} = \frac{\mu_o M}{4\pi R^3} \sqrt{1+3\mathrm{Sin}^2\varphi} \tag{2.10}$$

The phase angle of the field, φ', is a function of the vector position, which, in the case of a circle, is simply φ. From Fig. (2.2A)

$$\mathrm{Tan}\varphi' = \frac{B_N}{B_T} = 2\mathrm{Tan}\varphi \tag{2.11}$$

The magnitude (Eq. 2.10) and differential phase angle $\Delta\varphi=\varphi'-\varphi$ (derived from Eq. 2.11) of the planar field caused by M_{xy} are plotted in Graph 2.2 as a function of φ.

Fig. (2.2) applies the ideas developed in this section to the case of the earth's dipole. The field lying in the orbit plane is completely described by Equations (2.10) and (2.11). Since the orbit plane is at the "magnetic equator" of the dipole directed along the z axis, the z component of the field is, by the use of Eq. (2.9):

$$\bar{B}_z = \frac{-\mu M_z}{4\pi R^3}\,\bar{k} \tag{2.12}$$

Since M_{xy} and M_z are functions of θ and ψ_m alone, ψ_m and φ suffice to specify the field at any point in orbit. M_z, M_{xy} and φ_m are plotted in Graph (2.3) as a function of ψ_m for an orbit inclination of $\theta = 99.07°$. These quantities are derived from the expression for $\bar{M}$ in Table (2.1).

Note that M_z has two zero crossings which correspond to the intersections of the orbit plane and the cone generated by the earth's magnetic dipole as the earth rotates. (See Fig. (2.2b) . At these two points in orbit there is no normal magnetic field component. Also note that by choosing appropriate φ angles on either side of one of these zero crossings, the average $\overline{B}_z$ field may be made equal to zero for a portion of the orbit trace.

3.1 EDDY CURRENT DAMPING AND DRAG

When a conducting body moves in a magnetic field, currents are induced in the body which result in a body dipole moment. This dipole moment reacts with the environmental magnetic field to produce a torque. In the case of spin stabilized satellites, a component of the torque acts opposite to the spin vector producing "eddy current" drag which decelerates the satellite. This phenomenon was considered and reported upon in detail for early satellites as noted in reference (1) through (8). Reference (1) is particularly thorough and recommended for the critical reader. It is the only rigorous contemporary treatment found by the authors.

In practice, satellites are too complex to permit a formal solution of the magnetodynamic force equations. Hence,one is driven to the procedure of deriving the force equations for simple geometries and using the results for "rule of thumb" judgments on more complex configurations. To illustrate this method,the torque on a spinning, cylindrical, closed-end conductor will be derived.

The geometry is as shown in Fig. 3.1. The $\bar{B}$ vector lies in the x-z plane and the spin vector $\bar{\omega}$ is aligned with the z axis. The radius of the cylinder is (a) and its length is (L). The surface conductivity is σ_s. The voltage induced in an element ($ad\theta$) wide and (L) long is then:

$$e(\theta) = B_{\perp} L\omega a \cos\theta \tag{3.1}$$

Solution of Laplace's equation for the end plate with the voltage given by Eq. (3.1) imposed upon the boundary shows that the current flows in strips as sketched in Fig. (3.1b). This justifies writing the resistance of a loop as:

$$R = \frac{1}{\sigma_s}\left(\frac{2L}{ad\theta} + \frac{4a\cos\theta}{a\cos\theta d\theta}\right) = \frac{2(L+2a)}{\sigma_s a d\theta} \tag{3.2}$$

The current flowing in a loop is:

$$i = \frac{2e}{R} = \frac{2B_{\perp} L\omega a\cos\theta\sigma_s a d\theta}{2\ (L+2a)} \tag{3.3}$$

which causes the incremental dipole moment:

$$d\bar{M} = iA_{loop}\bar{j} = \frac{2L^2a^3\omega\sigma_s B_{\perp}\cos^2\theta d\theta}{(L+2a)}\bar{j} \tag{3.4}$$

The total dipole moment for the entire body is then:

$$\bar{M} = 2\int_0^{\pi/2} d\bar{M} = \frac{4L^2a^3\omega\sigma_s B_{\perp}}{(L+2a)}\int_0^{\pi/2}\cos^2\theta d\theta\,\bar{j}$$

$$= \frac{\pi L^2a^3\sigma_s\omega B_{\perp}}{(L+2a)}\bar{j} = \mu\omega B_{\perp}\bar{j} \tag{3.5}$$

Where

$$\mu = \frac{\pi L^2a^3\sigma_s}{L+2a}$$

This dipole moment interacts with the entire environmental field so that the total torque on the satellites is:

$$\begin{aligned} \bar{T} &= \bar{M}\times\bar{B}=\mu\omega B_\perp \bar{j}\times(B_\perp \bar{i}+B_{||}\ \bar{k}) \\ &= \mu\omega(B_\perp B_{||}\ \bar{i}-B_\perp^2\ \bar{k})=\bar{T}_p+\bar{T}_d \end{aligned} \tag{3.6}$$

Where

$$\bar{T}_p = \mu\omega B_\perp B_{||}\ \bar{i} \qquad \text{(Precessional Torque)}$$

$$\bar{T}_d = -\mu\omega B_\perp^2 \bar{k} = -\beta\omega\bar{k} \quad \text{(Drag Torque)}$$

$\beta = \mu B_\perp^2$ is an equivalent viscous drag coefficient. $\bar{T}_p$ and $\bar{T}_d$ are illustrated in Fig (3.1c). Observe that the torque, $\bar{T}$, is perpendicular to $\bar{B}$ and that $\bar{\omega}$, $\bar{B}$, and $\bar{T}$ are coplanar.

This analysis has ignored self and mutual inductance effects in the loops. The justification is the assumption of low spin rates. The value of μ computed in equation (3.5), when simplified, agrees with Smythe's Equation (9), but is in disagreement with Hooper (3). Review of Hooper's computation of the loop resistance suggests that the difference lies therein. Experimental work by Eide (Ref. 10) indicates that Smythe's Equation, uncorrected for end plate resistance, gives too high a drag torque. With the correction indicated in Eq. (3.5), agreement is improved.

The drag torque, $\bar{T}_d = -\beta\omega\bar{k}$, will cause the spin to decay exponentially. This is shown by solving the equation of motion:

$$J_s\, d\omega/dt = -\beta\omega \tag{3.7}$$

which has the solution:

$$\omega = \omega_o \epsilon^{-t/\tau} \tag{3.8}$$

Where $$\tau = \frac{J_s}{\beta} = \frac{J_s}{\mu B_\perp^2} = \frac{(L+2a)J_s}{\pi L^2 a^3 \sigma_s B_\perp^2}$$

Several investigators attempted to deduce the value of $\bar{B}$ by observing the decay time τ of early satellites. These efforts were frustrated by the complexity of the actual geometry. In practice, the value of the analysis is to show the relationship between the parameters so that they may be chosen to minimize drag or, conversely, to maximize it if damping is desired. A knowledge of the drag coefficient is desirable for choosing the number of spin-up rockets to be placed on a long life, spin stabilized satellite.

The same procedures used to reduce eddy current losses in motors and transformers are applicable to satellites. In general, however, eddy current drag is usually only a "phenomenon" and not a "problem" so that elaborate minimization efforts are not justified. Another approach to compensating for eddy current drag is an active system proposed by Grasshoff (11).

Graph (3.1) records the spin decay history of several spin stabilized satellites approximately 25 inches in diameter by 18 inches high which are densely "packed" with electronics and electro-mechanical devices. The satellites are in a nominally sun-synchronous, 500 n.m., circular orbit. Because of the polar inclination, the drag experienced indicates the maximum to be expected of satellites of this shape at this altitude. By scaling the field as a function of attitude, the results may be generalized.

Using the simplified "gyroscope" equation, it is easy to show that $\overline{T}_p$ tends to precess $\overline{\omega}$ into coincidence with $\overline{B}$. This suggests that some beneficial use might be made of the precessional torque for passive attitude control. One application which comes to mind is in a spin stabilized synchronous orbit satellite such as proposed by Williams for the first "Syncom" satellite (12). Since the earth's field is essentially tangential at the equator, the spin axis of the satellite will seek to align itself parallel to the earth's spin axis which is the desired operational orientation. Because the earth's field is small at this altitude, intentionally high damping coefficients would have to be designed into the satellite.

Stephen Marx of Philco's WDL Division recently pointed out that for a given volume and mass devoted to damping material, the most efficient damping is accomplished by a single turn coil, or coating, as opposed to an N turn coil (See Appendix C).

4.1 HYSTERESIS DAMPING AND RESIDUAL DIPOLE MOMENTS

It is a familiar result in the study of ferromagnetic materials that energy is expended when a sample of magnetic material is subjected to a cyclic magnetizing force. This energy is consumed in the process of mechanically rearranging the small magnetic domains composing the bulk material. These domains are small permanent magnets. If the domains are disarrayed in such a manner that their collective fields at a distance are negligible or undetectable, the sample is said to be demagnetized. On the other hand, if many domains are similarly aligned, the material possess a net dipole moment and will experience a torque when placed in a magnetic field.

Both of these effects have been used on satellites. Transit satellites 1B and 2A used hysteresis damping to despin the satellite to virtually zero RPM and then permanent magnets caused the satellite to align itself with the earth's magnetic field (1). On many other satellites, efforts have been taken to minimize the use of hysteresis materials. Unavoidable residual dipole moments are then carefully cancelled by a programmable"bias" coil on the vehicle after the drift due to the residual dipole has been measured.

Since magnetic attitude control, discussed in Section 7, includes the effect of magnetic dipoles, we shall restrict further discussion to hysteresis damping.

Let a Volume (V) of ferromagnetic material be rotated through 360 degrees at an infinitely slow rate in a uniform transverse magnetic field H. The work W (H) expended is a function of the strength of the field and the material itself. If this same material is placed in a spin stabilized satellite and subjected to a transverse magnetic field, the instantaneous rate of energy dissipated in the material due to hysteresis alone is:

$$p = W(H)f_s = \frac{W(H)\omega_s}{2\pi} \qquad (4.1)$$

Where ω_s = angular spin velocity (rad/sec)

f_s = spin rate (rev/sec)

An equivalent hysteresis drag torque can be postulated so that Eq. (4.1) can be written:

$$p = \omega T_h(H) \qquad (4.2)$$

Where $$T_h(H) = \frac{W(H)}{2\pi}$$

This torque is analogous to a kinetic friction which is zero at zero speed and a constant $W(H)/2\pi$ at any other speed. For oscillatory motion, it causes the equation to be non-linear. In the absence of other torques, the hysteresis drag torque will cause the satellite's spin rate to decay as shown by solving the equation of motion:

$$J_s \frac{d\omega_s}{dt} = -T_h(H) \tag{4.3}$$

To give: $$\omega_s = \omega_o - \frac{T_h(H)}{J_s} t \tag{4.4}$$

Unlike eddy current damping, hysteresis damping will cause the spin rate to reach zero in a finite time (T_o) given by:

$$T_o = \frac{\omega_o J_s}{T_h(H)} = \frac{\omega_o J_s 2\pi}{W(H)} \tag{4.5}$$

When both hysteresis and eddy current damping are present, the equation of motion is, by the use of Eq. (3.6)

$$J_s \frac{d\omega_s}{dt} = -\beta\omega - T_h(H) \tag{4.6}$$

Which has the solution:

$$\omega_s = \left[\omega_o + \frac{T_h(H)}{\beta}\right] \epsilon^{-\beta/J_s t} - \frac{T_h(H)}{\beta} \tag{4.7}$$

The time for ω_s to reach zero is:

$$T_o = \tau \ln\left(1 + \beta\omega_o / T_h(H)\right)$$

where τ is the damping time constant for eddy current damping alone. (See Eq 3.8)

Equation (4.7) can be rewritten in the form:

$$\omega_s = \omega_a \epsilon^{-n/N} - \omega_b \tag{4.9}$$

Where (n) is orbital passes in hundreds, (N) is the eddy current damping time in hundreds of Revs., and ω_s, ω_a and ω_b are in RPM. By manipulating Eq (4.9), N, ω_a and ω_b may be solved for and evaluated numerically from the data given in Graph 3.1.

Letting $\omega_o = \omega_s(0)$, $\omega_1 = \omega_s(n_1)$, $\omega_2 = \omega_s(n_2)$ and setting $n_2 = 2n_1$ then:

$$\omega_b = \frac{\omega_1^2 - \omega_o\omega_2}{\omega_o + \omega_2 - 2\omega_1} \tag{4.10}$$

$$\omega_a = \omega_o + \omega_b \tag{4.11}$$

$$N = \frac{n_2}{\ln\left(\frac{\omega_a}{\omega_2+\omega_b}\right)} = \frac{n_2}{\ln\left(\frac{\omega_o+\omega_b}{\omega_2+\omega_b}\right)} \tag{4.12}$$

From Graph (3.1) with the zero reference for this computation shifted to n = 2 (200 revs): $\omega_o(0) = 13.89$, $\omega_1(100) = 13.33$, $\omega_2(200) = 12.79$, which yields the result: $\omega_b = 1.79$, $\omega_a = 15.68$, N = 27.66, therefore: $\omega_s = 15.68\ \epsilon^{-n/27.66} - 1.79$ (4.13)

This result indicates that both eddy current and hysteresis damping are significant for long lifetime satellites. Neither type of damping is intentionally incorporated into the design of the spacecraft described in Section 3.

For intentional hysteresis damper design information the reader is directed to Refs (1) and (2), and Appendix C.

5.1 D.C. SPIN CONTROL MOTOR

In the previous sections, the drag torques acting to slow a spin stabilized satellite have been discussed and their magnitude recorded from actual flight tests. This section describes two circuits currently being built and scheduled to fly soon which overcome these drag torques and permits essentially constant satellite spin rates. As a result of constant spin rate, other systems in the spacecraft have been greatly simplified.

The D.C. spin control motor utilizes the planar field encountered in orbit as a "stator" field. The satellite is made the armature with its spin axis normal to the orbit plane. A multi-turn coil with its normal perpendicular to the spin axis (hence in the orbit plane) generates a dipole moment which interacts with the earth's magnetic field. In Section2 it was pointed out that the normal field to the orbit plane was zero twice per orbit. Operation of the motor in the vicinity of these orbital positions results in acceleration about the spin axis only. That is, operation does not cause a precessional torque. Commutation of the armature coil is accomplished in either of two ways: (1) with one horizon sensor and a monostable flip-flop or (2) with two horizon sensors. These sensors are mounted perpendicular to the spin axis and phased with respect to the armature coil normal as dictated by orbit altitude and planned geographical point of operation. With the control station in the United States, it is easy to command the satellite near the point where the earth's dipole lies entirely in the orbit plane so the horizon sensors are placed accordingly. Another advantage of motor operation in this region

is that it corresponds to the orbit "north magnetic pole" where the field strength is twice that at the orbit's "magnetic equator."

Fig. 5.1 illustrates the placement of the coil and horizon sensors on the spacecraft. The angle β is computed from the formula:

$$\beta = \mathrm{Sin}^{-1}(1 + {}^{h}/_{R_e})^{-1} \qquad (5.1)$$

where: h = orbit altitude and R_e = earth radius (3444 NM). For h = 500 NM, the angle β is approximately 61°. A plot of β as a function of (h) would reveal that commutation is virtually immune to altitude injection errors. ($\pm\Delta\beta = 1^\circ$ corresponds to ± 50 NM injection errors)

The idealized logic for achieving commutation is depicted in Fig. (5.2). Either arrangement (1) or (2) may be utilized to achieve commutation. With additional logic circuits, arrangements (1) and (2) may both be used in the same satellite, arrangement (1) being a "back-up" mode of operation to arrangement (2).

Referring to Fig. (5.1) and (5.2), the operation of the motor is as follows: Horizon sensor HS_1 crosses the earth's horizon and generates a spike signal which triggers F.F. #1 to such a state that the dipole moment $\overline{M}_a$ is directed opposite to the vector normal $\overline{n}$ of the coil. From the formula:

$$\overline{T} = \overline{M}_a \times \overline{B} = \overline{M}_a B \mathrm{Sin}\xi \overline{S}_u \qquad (5.2)$$

A torque tending to accelerate the satellite is developed. As the satellite makes one half a revolution, ξ will vary from 180° to 0°. At $\xi = 0^\circ$, horizon sensor #2 will cross the earth's horizon and generate a trigger signal which reverts the flip-flop #1 to its initial state which in turn causes the current in the coil to generate a dipole moment aligned with the normal $\overline{n}$. Again the angle between $\overline{M}_a$ and $\overline{B}$ will vary between

180° and 0° so that an accelerating torque is again developed.

If deceleration had been desired, the polarity of the current in the coil would have been reversed with the "spin-up, spin-down switch."

Operation with a single horizon sensor and monostable flip-flop would be quite similar except that T/2 seconds after HS #1 generated a spike, the monostable flip-flop would generate the reset pulse instead of HS#2.

5.2 TORQUE DEGRADATION DUE TO NON-OPTIMUM COMMUTATION ANGLE

Since the motor commutates from a horizon pulse, there can be only one place in the satellite's orbit where optimum commutation will occur. That position is when $\varphi = \varphi_m$. To calculate the degradation in torque when the motor is operated at other orbital positions, assume that the angle between $\bar{B}$ and $\bar{M}_a$ when commutation occurs is δ degrees. (See fig. 5.3). The average torque on the satellite is then:

$$\tilde{\bar{T}} = \frac{2MB}{\pi} \quad \text{Cos}\delta \bar{S}_u \tag{5.3}$$

Recalling the definition φ' from Table (2.1), Eq. (5.3) can be rewritten in terms of φ' as:

$$\tilde{T} = \frac{2MB}{\pi} \text{Sin}\varphi' \tag{5.4}$$

From Graph (2.1) it will be recalled that $\bar{B}_{xy}$ tends to "pull" towards $\bar{M}_{xy}$ in the vicinity of $\varphi = \varphi_m$. Therefore, within $\pm 10°$ of $\varphi = \varphi_m$, φ' varies by only $\pm 5°$ and the torque as given by Eq. (5.4) does not vary appreciably with satellite position near $\varphi = \varphi_m$.

5.3 TORQUE DEGRADATION DUE TO CLOCK COMMUTATION

If commutation is accomplished using one horizon sensor and a timing circuit, the torque becomes a function of the satellite spin rate. Let the design nominal spin rate be:

$$\omega_o = 2\pi f_o = \frac{2\pi}{T_o} \tag{5.5}$$

The commutation angle δ is $\frac{\omega T_o}{2}$. The average torque $\tilde{T}$ as a function of the ratio ω/ω_o is: (See Fig. 5.4).

$$\tilde{T} = \frac{MB}{\pi}(1 - \cos\delta) = \frac{MB}{\pi}(1 - \cos\pi\frac{\omega}{\omega_o}) \tag{5.6}$$

Should the satellite be injected at too high a spin rate, natural drag torques will despin the satellite down through the higher zero's of the torque-speed curve. The zero at $\omega/\omega_o=0$, however, emphasizes that the motor has no starting torque.

5.4 NOMINAL ACCELERATION AND SPIN GRANULARITY

For the two scanner commutator, the equation of motion is:

$$\tilde{T} = \frac{2MB}{\pi}\sin\varphi' \approx \frac{2MB}{\pi} = J_s\frac{d\omega}{dt} \tag{5.7}$$

Integrating and substituting the following assumed parameter values:

$B = 4.8\times10^{-5}$ Weber/Meter^2

$M = 1.5$ Amp-Meter^2 (Coil Dipole Moment)

$J_s = 2$ Kg-Mt^2 (Satellite Moment of Inertia about Spin Axis)

T = time of motor operation in Sec.

$$\Delta\omega=\omega_2-\omega_1=\frac{2MBT}{\pi J_s}=2.29\times10^{-5}T\ \frac{Rad}{Sec} \tag{5.8}$$

$$=2.19\times10^{-4}T(RPM)=3.76\times10^{-3}\Delta\varphi(RPM)$$

If the motor is operated for say 20 degrees of the orbit, $\Delta\varphi = 20^\circ$ so that: $\Delta\omega = 7.52\times10^{-2}$ RPM. If the motor operates for a fixed period of time after initial commanding, then Eq. (5.8) fixes the minimum spin-up or spin-down increment for the motor. If the particular application does not require extremely accurate speed control, the fixed time of operation circuit is easily and reliably implemented with a minimum power load.

5.5 FLUX METER

When the motor is not in operation, the voltage induced in the coil can be telemetered which will reveal the magnitude of the planar component of the earth's field.

$$v_a(\varphi,t)=B_{xy}(\varphi)NA\omega\cos\omega t \tag{5.9}$$

Which yields $$B_{xy}(\varphi)=\frac{NA\omega}{\sqrt{2}V_a(\varphi)}$$

Where $V_a(\varphi)$ is the RMS value of the telemetered signal as a function of orbit position.

5.6 DESATURATION OF GYROS

The ideas discussed in Sect. (5.1) through (5.4) are directly applicable to the despinning of saturated gyros. In Sect. (8.1) the authors propose a three axis attitude control scheme utilizing a single induction motor (or gyro) aligned with the orbit normal to maintain earth centering of a satellite payload. As the angular momentum of the motor changes during an orbit because of accumulated perturbing torques,

the rotor of the motor will have to be spun up or down (desaturated). In this satellite configuration, the torque coil normal is directed along the satellite velocity vector so that polar operation results in a torque of:

$$\bar{T} = \pm MB\bar{S}_u \text{ Nt-Meters} \tag{5.10}$$

and the momentum that is "dumped" in the time T is:

$$\Delta\bar{H} = \Delta(J_s\bar{\omega}_s) = \bar{T}T = \pm MBT\bar{S}_u \tag{5.11}$$

6.1 SYNCHRONOUS MOTOR SATELLITE

In search for a simple, closed-loop attitude control scheme utilizing the principles developed in Section 7.1, the authors were intrigued by the ability of a synchronous motor satellite to determine its orbital position. The system compares a horizon pulse with the phase of the motor signal and computes the orbital latitude from the difference. A suggested reference for another somewhat related concept is Ref (1).

Another application of the Synchronous Motor Satellite is its use as a synchronous modulator. A transducer on the surface of the satellite, measuring the Sun's flux for example, will have its output modulated by the spin of the satellite. Since the spin rate is identical to the frequency of the on board oscillator, the output of the transducer can be synchronously demodulated on board the spacecraft or both the modulated carrier and reference signal can be transmitted to the ground for subsequent demodulation.

The scheme although apparently promising, has substantial difficulties. The equations of motion of the satellite, assuming only natural eddy current damping, has virtually no damping and an extremely long period. The current progress in dampers for use on gravity gradient satellites suggests that the synchronous motor satellite will possibly be practical someday though it admittedly is only an idea at the present time. (2)(3)

The coordinate system and conventions used to derive the equation of motion of the satellite is shown in Fig. 6.1 and Table 6.1. The damper is shown as a simple viscous drag. The objective of the analysis is to obtain the "power angle" of the motor as a function of the displacement of the magnetic field.

To simplify the handling of the initial conditions, a new set of variables are introduced:

$$\begin{array}{lll} \theta_m = \omega_c t & \theta_m(0) = 0 & \dot{\theta}_m(0) = \omega_c \\ \theta_s = \xi_s - \omega_c t & \theta_s(0) = 0 & \dot{\theta}_s(0) = -\omega_c \\ \theta_d = \xi_d - \omega_c t & \theta_d(0) = 0 & \dot{\theta}_d(0) = -\omega_c \end{array} \qquad (6.1)$$

The power angle, θ_{pa} is defined as $\theta_B-(\theta_s+\theta_m) = \theta_B-\xi_s$

Writing the force equations for the satellite frame, the damper, the linearized motor torque equation, etc., and solving for the Laplace Transformed power angle:

$$\theta_{pa} = \frac{S^2[S+a_o]\theta_B(S)}{S^3+aS^2+bS+c} \qquad (6.2)$$

Where:

$$a_o = a = \frac{J_o+J_s}{J_oJ_s}\beta$$

$$b = \frac{MB}{J_s}$$

$$c = \frac{MB\beta}{J_dJ_s}$$

Because of the small magnitude of magnetic field B the system is still underdamped for realistic values of M and β. Thus the synchronous motor satellite must await further viscous damper development to be practicable.

7.1 TWO DEGREE ATTITUDE CONTROL (SPIN STABILIZED)

The torques caused by the earth's field acting upon permeable and conducting satellite construction materials were not fully accounted for in early satellites and resulted in unexpected attitude drifts.(1)(2) Though the effects of Eddy Current Torques were well documented in the literature (3)(4)(5), the effects of residual magnetic dipoles seems to have been overlooked. Studies were initiated aimed at nulling out the satellite residual dipole moment which causes precessional torques with a satellite bias coil. Subsequent consideration of the force equations describing the interactions of these two fields resulted in a number of proposals advocating the use of the earth's field in attitude control schemes. Probably the first was a proposal by L. H. Grasshoff at RCA's Astro-Electronics Division.(6) A less sophisticated mechanization of his ideas has been incorporated on the Tiros series of satellites from Tiros II on. Other schemes are recorded in the references (7)(8)(9)(10)(11)(12).

The two degree control system utilized in the program is an out growth of Grasshoff's and others' work at RCA-AED and is credited to them. As will be seen, the theory of operation of the system is very simple and the hardware reliability has in actual flight tests proven itself 100% reliable over a period of more than two years. Agreement between actual and predicted system behavior has been identical within

the limit of attitude sensing accuracy. (An outline of the attitude sensing system is given in Appendix A.)

The attitude control system consists of an N turn coil, with an area of (A) Meters^2 with the normal to the plane of the coil coinciding with the satellite spin axis. The coil is excited with a current of (I) amperes. The polarity or sense of the current through the coil is reversed four times per orbit. The reversal is accomplished by a clock filling a register, the capacity of which is preset on the ground to correspond to a nominal orbit. The "phasing" or initial start time for the square wave current signal is commanded by a telemetered time delay code sent to the satellite upon ground station passage.

The current in the coil causes a dipole moment M=NIA which in vector form is:

$$\overline{M} = \overline{0}, +M_s\overline{S}_u, \text{ or } -M_s\overline{S}_u \qquad (7.1)$$

(Coil design and power requirements are discussed in Appendix B) With the aid of Fig. 7.1, Equations (2.2) or (2.7) and the simplified "Gyroscope Equation":

$$\overline{T} = \overline{\omega}_p \times \overline{H} = \overline{\omega}_p \times J_s \omega_s \overline{S}_u \qquad (7.2)$$

the precession vector $\overline{\omega}_p$ will be derived. The results of the analysis will show that $\overline{\omega}_p$ lies in the orbit plane (xy plane) and its angular position, φ_p is a single valued function of the coil current phasing angle φ_t. Thus φ_t completely determines the axis of precession and the number of torquing cycles determines the magnitude or total arc through

which $\bar{S}$ is precessed. $\bar{S}$ may be pointed in any direction, though normal mission attitude has it parallel to the z axis.

For clarity, $\bar{\omega}_p$ will be derived using the simplified model of the earth's field given by Eq. (2.7). The torque on the satellite dipole moment $\bar{M}_s$ is $\bar{M}_s \times \bar{B}$ which is set equal to $\bar{T}$ in Eq. (7.2)

$$\bar{T} = \bar{M}_s \times \bar{B} = M_s(t)\bar{S}_u \times \bar{B} = \bar{\omega}_p \times J_s \omega_s \bar{S}_u \qquad (7.3)$$

Where

$\bar{B}$ = instantaneous magnetic induction of Earth's field in Weber/Mt.^{2}

$\bar{M}_s$= instantaneous dipole moment of satellite in Amp-Meter^2

J_s= Spin axis moment of inertia in Kg-Meters^2 (also the max. principal moment)

$\bar{S}$ = Spin axis (assumed coincidental with the axis along which J is measured) and has the magnitude of rad/sec.

$\bar{S}_u$= Unit vector parallel to S.

$\bar{\omega}_p$= Gyroscopic precession vector.

Rearranging Eq. (7.3);

$$\bar{S}_u \times (M_s(t)\bar{B} + J_s \omega_s \bar{\omega}_p) = \bar{0} \qquad (7.4)$$

Which has the solution:

$$\bar{\omega}_p = - \frac{M_s(t)\bar{B}}{J_s \omega_s} \qquad (7.5)$$

The average value of $\bar{\omega}_p$ is:

$$\tilde{\bar{\omega}}_p = \frac{1}{2\pi}\int_{\varphi_1}^{\varphi_1+2\pi} \bar{\omega}_p d\varphi = \frac{-1}{2\pi}\int_{\varphi_1}^{\varphi_1+2\pi} \frac{M(t)}{J_s\omega_s}\,\bar{B}\,d\varphi \tag{7.6}$$

Substituting for $\bar{B}$ from Eq. (2.7), the value of $\tilde{\bar{\omega}}_p$ is:

$$\tilde{\bar{\omega}}_p = \frac{3\mu_o M_e M_s \sin\theta}{4\pi^2 R^3 J_s \omega_s}\left(\cos 2\varphi_1\,\bar{i} + \sin 2\varphi_1\,\bar{j} + 0\bar{k}\right) \tag{7.7}$$

Eq. (7.7) displays that $\bar{\omega}_p$ lies in the orbit plane and justifies the complex representation of $\bar{\omega}_p$ as:

$$\tilde{\bar{\omega}}_p = \omega_p \varepsilon^{i\varphi_p} \tag{7.8}$$

Where

$$\omega_p = \frac{3\mu_o M_e M_s \sin\theta}{4\pi^2 R^3 J_s \omega_s}$$

And

$$\tan\varphi_p = \frac{\sin 2\varphi_1}{\cos 2\varphi_1}$$

$$\varphi_p = 2\varphi_1 \tag{7.9}$$

If the more accurate expression, Eq. 2.2 had been used to represent the earth's magnetic field in integrating Eq. (7.7), a very complicated expression for $\tilde{\bar{\omega}}_p$ would have resulted. The accuracy gained is more than negated by the difficulty in readily interpreting the results. One question which does need to be answered through is whether or not a (z) component of precession results. The answer is in the affirmative, but its magnitude is three orders of magnitude down from the planar component for the example orbit assumed throughout the paper.

One correction must be applied to the phasing of the coil current φ_t, however. Eq. (7.6) illustrates this rather clearly. There is no "cross-coupling" in the relationship of the components of $\bar{\omega}_p$ and $\bar{B}$. The planar field components give rise to the planar precession vector. The phasing φ_t in Eq. (7.9) is based upon the idealization that $\bar{M}_{xy} = M_y\bar{j}$. In reality, of course, M_{xy} may either rotate in the x-y plane or oscillate back and forth depending upon the orbit inclination. Therefore, the phasing angle φ_t should be measured from the "mean" magnetic equator for an orbital pass. This is just $(\varphi_m - \pi/2)$ where φ_m is given in Table 2.1 and illustrated in Graph 2.3 for the example orbit. It is convenient to compute commands based upon the orbit ascending node so an expression for a new phasing angle will be derived. Referring to Fig. 7.2,

φ_p = Precession vector based on orbit magnetic equator

φ_{pan} = Desired precession vector based on ascending node

φ_t = Current phasing angle based upon the orbit "magnetic equator."

φ_{tan} = Current phasing angle based upon orbit ascending node.

φ_m = Angular position of M_{xy} (ψ_m)

By Eq. (7.9) $\varphi_p = 2\varphi_t$ and from the figure

$$\varphi_{tan} = \varphi_m + \varphi_t - \pi/2 \tag{7.10}$$

$$\varphi_{pan} = \varphi_p + \varphi_m - \pi/2 \tag{7.11}$$

Solving for φ_{tan} in terms of φ_{pan}

$$\varphi_{tan} = \tfrac{1}{2}(\varphi_{pan} + \varphi_m) - \frac{\pi}{4} \tag{7.12}$$

7.2 ANGULAR PRECESSION FOR HALF ORBIT OPERATION

From Eq. (7.8), the angle Ω through which the spin axis is precessed in one half orbital period is:

$$\Omega = \int_0^{T_o/2} \omega_p dt = \frac{3\mu_o M_e M_s \sin\theta T_o}{8\pi^2 R^3 J_s \omega_s} \qquad (7.13)$$

From Appendix B, $M_s = 1.75$ Amp-Meter^2

From Graph 2.3, $\frac{\mu_o M_e}{4\pi R^3} = 2.4 \times 10^{-5}$ Webers/Meter^2

And from the satellite geometry and nominal operation:

$J_s = 2.26$ Kg-Meter^2

$\omega_s = 1.57$ Rad/Sec (15 RPM)

$T_o = 6{,}180$ Sec (103 Minutes)

$\sin\theta = 0.9875$

Substituting numerical values, Ω_p is approximately 2°/half orbit. The granularity in the attitude control system is one-half Ω_p or $\pm 1.0^\circ$ in the present case. If finer control is required, Ω_p could be made smaller by reducing the magnitude of the coil dipole moment M_s. The magnitude of the coil current could be made ground programmable if various torquing rates are desired. Experimentally, the observed precession rates of the spin axis have coincided with those predicted by Eq. (7.13).

8.1 THREE DEGREE ATTITUDE CONTROL (ONE INERTIA STORAGE ELEMENT)

In survey articles on the "state of the art" and "capabilities of various attitude control systems, magnetic attitude control schemes

are, in general, summarily dismissed. In part, this is due to the complexity of the stereotyped "general" magnetic attitude control system composed of three torque coils, three flux meters, three inertia storage wheels, and a fairly sophisticated on-board computer. It is, of course, unfair to saddle the name or general concept with such an arbitrary mechanization. The configuration proposed by the authors in this section is a "three axis magnetic attitude control system" but certainly simpler than the "general" system mentioned above. Hopefully, it will provide an alternative for analysis and discussion on the capabilities of various near-earth satellite attitude control systems.

The proposed three axis system evolves from the two degree system described in Section (7.1) without major modification of the theory presented there. Instead of spinning the entire satellite to obtain inertial rigidity, a motor running at high speed, is used to provide the angular momentum required for "gyroscopic stiffness". The axis of the motor is aligned so as to coincide with the spin axis discussed in Section (7.1). The third degree of freedom is controlled by varing the speed of the motor about its quiescent value.

As in any reaction wheel control device, perturbance torques cause the wheel to accelerate or decelerate so as to produce an equal and opposite force. Most of the perturbance torques are periodic with respect to the orbit period and have an average value of zero. Thus, the storage system should have a capacity to permit time averaging over several orbital periods without saturating. The long term net momentum change

is dumped as described in Section (5.6). Unlike dumping against the torque of a gas jet, dumping against the earth's magnetic field may be accomplished at a leisurely rate which relaxes the maximum torque requirements on the motor.

Although a number of conventional servomechanism circuits could be employed to control the reaction wheel, the authors would like to propose a mechanization that they have not seen elsewhere. Based upon previous experiences, D. C. motors are dismissed out of hand. The alternative "run-of-the-mill" A.C. servomotor systems are equally unattractive because of their inefficiency in the momentum storage mode of operation. As a result, the authors considered the use of a "power" type induction motor instead of a "servomotor" design. Since the motor is to be run near synchronous speed, the efficiency is excellent. The motor itself is run in a "balanced" mode so that the usual myraid of servomotor design considerations over single phasing, etc. are avoided, contributing additional efficiency. Lastly, by varing the frequency of the power supply instead of the amplitude of the voltage applied to the "control" winding, the proposed system dispenses with a control-winding servo-amplifier.

Though a detailed analysis of the control system and motor design considerations are too volumnous for inclusion here, an idealized sketch will be given. The overall system is pictured in Figure (8.1). The system block diagram is shown in Figure (8.2) and a detailed angle convention listing is presented in Figure (8.3) and Table 8.1.

With balanced voltages, the torque slip equation of an idealized induction motor is (1):

$$T = \frac{\pi \varphi_m^2 \omega_s R}{2\,[R^2+(L\omega_s \pi/2)^2]} \tag{8.1}$$

Where φ_m = rotating flux (Webers)

R = amature resistance (Ohms)

L = armature inductance (Henrys)

ω_s = slip angular frequency (Rad/Sec.)

Rather than making R large, as is the case in servomotor design, let R be small as in "power" induction motor design. Then, assuming operation near synchronous speed, the simplified torque equation becomes:

$$T = \frac{\pi \varphi_m^2}{2R}\ \omega_s = k_t \omega_s \tag{8.2}$$

Where k_t is quite large due to R being small.

In terms of the angles defined in Figure (8.3), Eq. (8.2) can be written:

$$T = k_t[\omega_a(1-k_p\varepsilon-k_i\int \varepsilon dt)+\dot{\theta}_s-\dot{\theta}_a] \tag{8.3}$$

The response of the system to an arbitrary requirement $\theta_r(t)$ and perturbation torque, $T_p(t)$ will now be derived. The assumed initial conditions are:

$$\theta_a(0) = \theta_s(0)=\dot{\theta}_s(0)=0$$
$$\dot{\theta}_a(0) = \frac{k_t\omega_a}{(\beta+k_t)}=\dot{\theta}_{ao} \tag{8.4}$$

The principle of conservation of angular momentum applied to the system results in the Laplace transformed equation:

$$J_a S^2\theta_a(S)-J_s S^2\epsilon(S)=T_p(S)+J_a\dot{\theta}_{ao}-J_s S^2\theta_r(S) \qquad (8.5)$$

Writing the force equation for the armature and performing the Laplace transformation yields:

$$S\left[J_a S+(\beta+k_t)\right]\theta_a(S)+\left[(\beta+k_t)S+k_t\omega_a k_p+k_t\omega_a k_p+k_t\omega_a k_i/S\right]\epsilon(S)$$
$$= J_a\dot{\theta}_{ao}+(\beta+k_t)S\theta_r(S)+k_t\omega_a/S \qquad (8.6)$$

Solving equations (8.5) and (8.6) for $\epsilon(S)$ yields:

$$\epsilon(S) = \frac{S^2\left[S+\frac{(J_a+J_s)}{J_a J_s}(\beta+k_t)\right]\theta_r(S) - \frac{1}{J_s}\left[S+\frac{(\beta+k_t)}{J_a}\right]T_p(S)}{S^3+\frac{(J_a+J_s)}{J_a J_s}(\beta+k_t)S^2+\frac{\omega_a k_t k_p}{J_s}S+\frac{k_t\omega_a k_i}{J_s}}$$

$$= \frac{S^2(S+a_o)}{S^3+aS^2+bS+c}\,\theta_r(S) - \frac{(S+b_o)/J_s}{S^3+aS^2+bS+c}\,T_p(S)$$

Where: $a = a_o = \frac{J_a+J_s}{J_a J_s}(\beta+k_t)$

$$b_o = (\beta+k_t)/J_a$$

$$b = \omega_a k_t k_p/J_s$$

$$c = \omega_a k_t k_i/J_s$$

Applying the Routh Stability criteria to the system, we have the only stability restraint on the parameters:

$$k_i < k_p(\beta+k_t)(J_a+J_s)/J_a J_s \cong k_p(\beta+k_t)/J_a \qquad (8.8)$$

The requirement variable, $\theta_r(t)$ is essentially a ramp function caused by the apparent rotation of the Earth around the satellite at the orbital

rate. The two most probable forms of the perturbation torques $T_p(t)$ on the satellite are impulse and periodic functions. All of these requirements are met by the control system. A step function perturbation torque will cause the integrator to saturate in time but it can be "reset" during the momentum dumping cycle if need be.

Assuming that the factors of the characteristic equation are a simple pole and complex conjugate poles, the response to a ramp function $\omega_o t=\theta_r(t)$ and an impulse perturbation is of the form:

$$\epsilon(t)=A\epsilon^{-\gamma t}+B\epsilon^{-\alpha t}\mathrm{Sin}(\beta t+\psi) \qquad (8.9)$$

Where A, B, γ, α, β and ψ depend upon the system parameters listed in Eq. (8.7). Reference (2) supplies detailed information on the Laplace Transform theory used in this section.

References (3), (4), (5), (6), (7), and (8) discuss the general problem of reaction wheel control devices and are recommended for a more detailed discussion of the topic.

It is of interest to compute the sizing of various components to convert the spin stabilized satellite of Section 7 to a three degree controlled system. The angular momentum assumed in Section 7 was:

$$\begin{aligned} J_s\omega_s &= 2.26\ \text{Kg-Meter}^2 \times 15\text{RPM} \\ &= 1.13\pi\ \text{Kg-Meter}^2\text{-Sec} \end{aligned} \qquad (8.10)$$

The armature (or attached fly wheel) of the induction motor with a synchronous speed of 6000 RPM should have an inertia of:

$$J_a=\frac{J_s\omega_s}{\omega_a}\quad\frac{1.13\pi}{2\pi\times100}=0.0377\ \text{Kg-Meter}^2 \qquad (8.11)$$

to duplicate the gyroscopic behavior of the spin stabilized satellite. Assuming a cyclic perturbation of 100 dyne centimeters, the maximum change in the stator field oscillator frequency (nominally 100 cps) during an orbit is:

$$\Delta f = \frac{T_p(T_o/2)}{2\pi J_a} = 0.26 \text{ CPS} \qquad (8.12)$$

The fractional deviation $\delta = \Delta f/f = 0.0026$ is virtually negligible so that constant amplitude control of the oscillator is easily achieved. The synchronous speed is chosen in the vicinity of 100 cps as opposed to 400 cps or more to minimize windage and magnetic losses.

TABLE 2.1

$\bar{X}\ \bar{Y}\ \bar{Z}$	Quasi-stationary primary reference frame. (The XY plane coincides with the Earth's equatorial plane. The X axis is displaced by the angle γ from Aries and corresponds to the line of nodes in an Euler system)
xyz	Orbital-reference frame (non-rotating). (The xy plane is the orbital plane and the z axis is the orbit normal in the right hand sense.)
θ	Orbit inclination
$\bar{R}$	Position vector of satellite or point of observation.
$\bar{R}_u$	Unit Vector derived from $\bar{R}$.
φ	Angle between $\bar{R}$ and X (or x) in xy plane.
θ_m	Inclination of earth's dipole moment (11.5°).
$\psi_m=\omega_m t$	Angle between the XY plane component of M and the X (or x) axis. Disregarding orbit regression, ω_m is $\pi/12$ radians per hour.
$(A_t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \text{Cos}\theta & \text{Sin}\theta \\ 0 & -\text{Sin}\theta & \text{Cos}\theta \end{bmatrix}$	Transformation matrix from XYZ coordinates into xzy coordinates.
$\bar{I}, \bar{J}, \bar{K}$	Unit vectors in the XYZ frame.
$\bar{i}, \bar{j}, \bar{k}$	Unit vectors in the xyz frame.
$\bar{R}_u$	$= [\text{Cos}\varphi\bar{I}+\text{Sin}\varphi\text{Cos}\theta\bar{J}+\text{Sin}\varphi\text{Sin}\theta\bar{K}]$
	$= [\text{Cos}\varphi\bar{i}+\text{Sin}\varphi\bar{j}]$

$\overline{M}$ $= M[\mathrm{Sin}\theta_m \mathrm{Cos}\psi_m \overline{I} + \mathrm{Sin}\theta_m \mathrm{Sin}\psi_m \overline{J} + \mathrm{Cos}\theta_m \overline{K}]$

$= M[\mathrm{Sin}\theta_m \mathrm{Cos}\psi_m \overline{i} + (\mathrm{Cos}\theta \mathrm{Sin}\theta_m \mathrm{Sin}\psi_m + \mathrm{Sin}\theta \mathrm{Cos}\theta_m)\overline{j}$

$+(-\mathrm{Sin}\theta \mathrm{Sin}\theta_m \mathrm{Sin}\psi_m + \mathrm{Cos}\theta \mathrm{Cos}\theta_m)\overline{k}]$

M_z Magnitude of the z component of the Earth's dipole moment.

Mxy Magnitude of the component of the Earth's dipole moment lying in the orbit (xy) plane.

B_n Flux density normal to the satellite velocity vector and lying in the plane; assumed circular orbit.

B_t Flux density directed opposite to the satellite velocity vector and lying in the plane; assumed circular orbit.

φ' Phase angle between B_n and B_t in the sense shown in Figs. (2.2) and (2.3).

φ_m The angle between Mxy and the x axis.

$\overline{S}$ Satellite spin vector in Rads./Sec.

$\overline{S}_u$ Unit vector along $\overline{S}$.

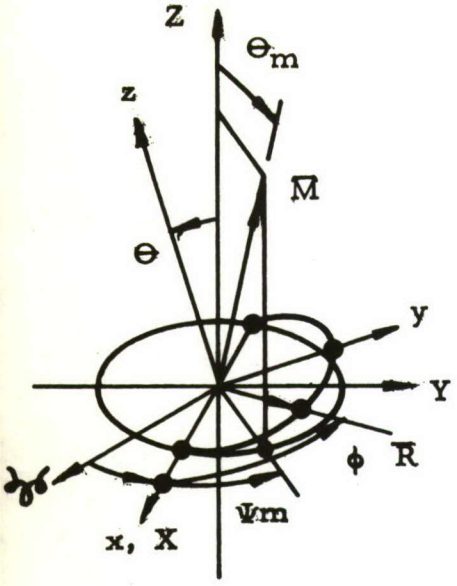


FIG. 2.1

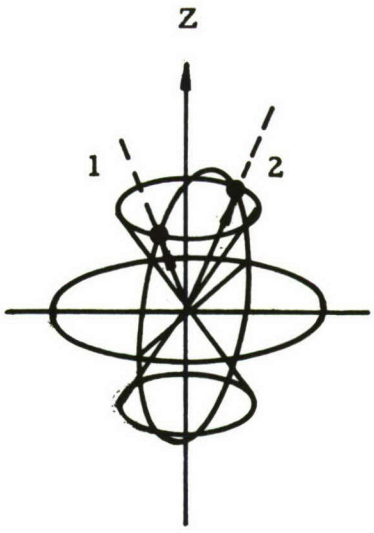


FIG. 2.2 (B)

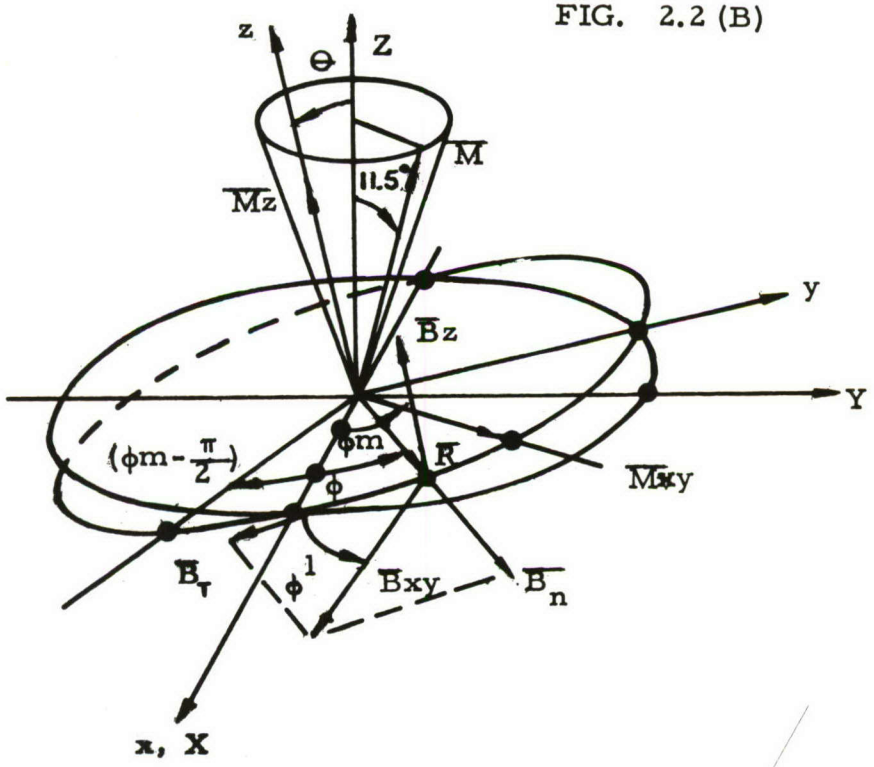
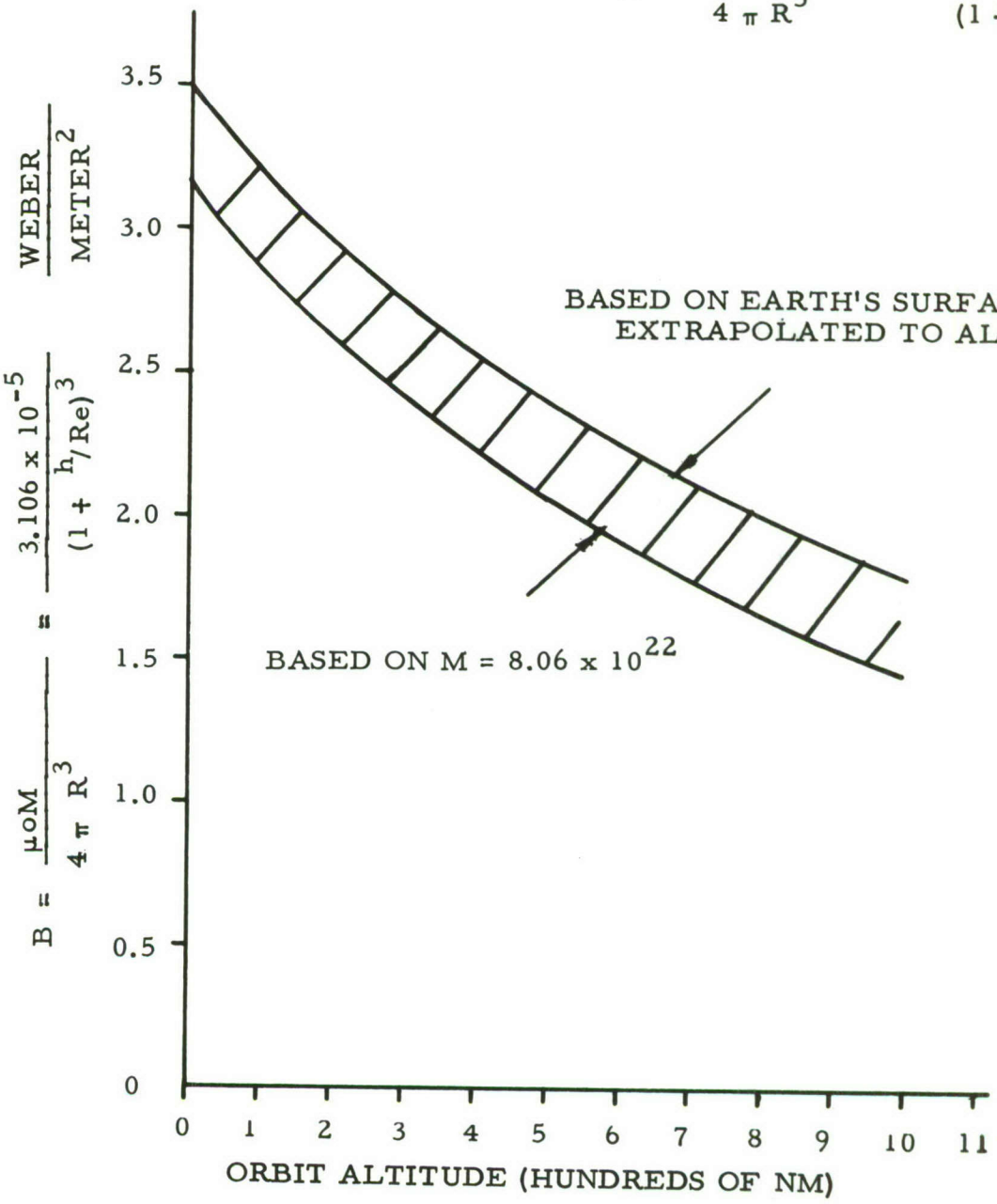


FIG. 2.2 (A)

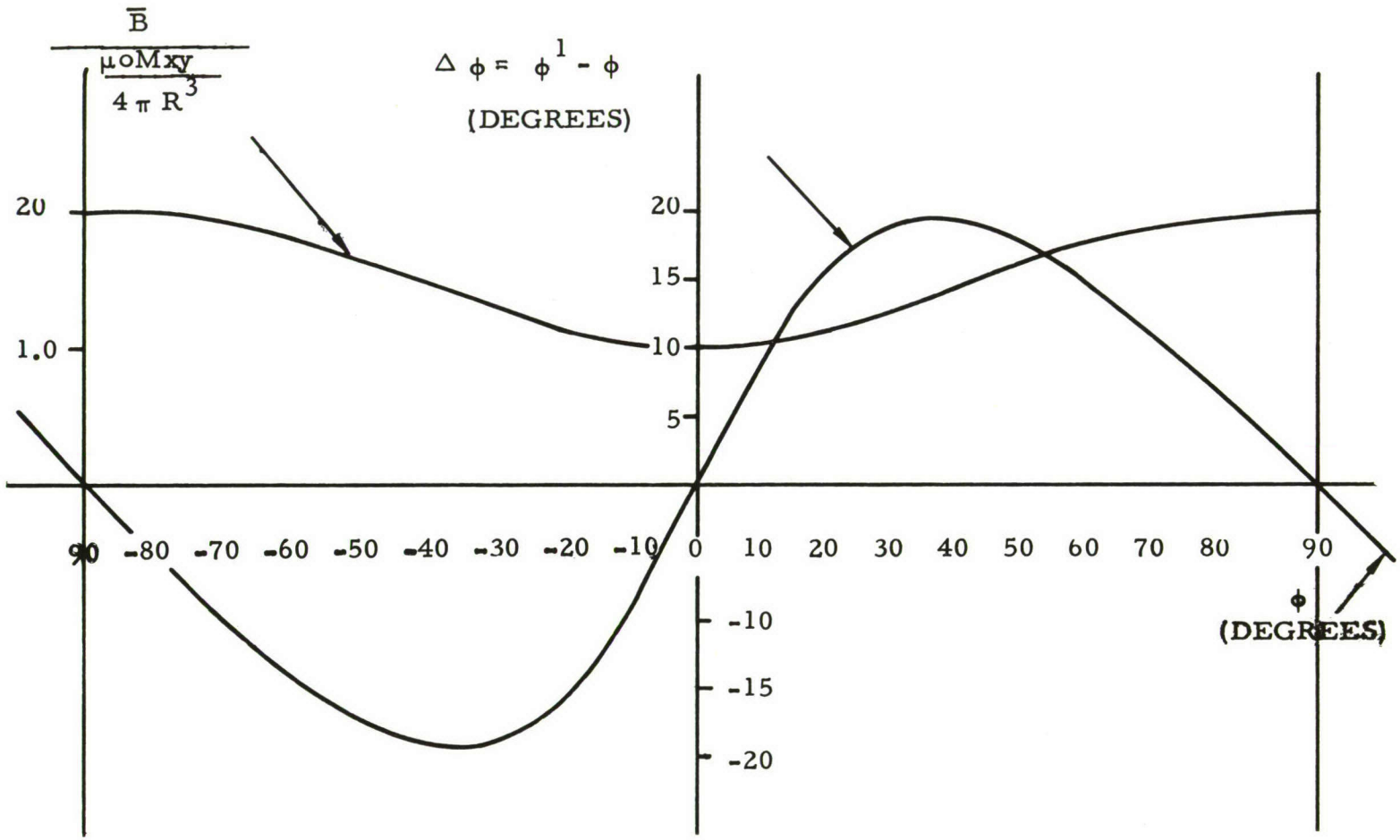
MAGNETIC FIELD VARIATION WITH ALTITUDE (NM)

$$B = \frac{\mu_o M}{4\pi R^3} = \frac{3.106 \times 10^{-5}}{(1 + h/Re)^3}$$



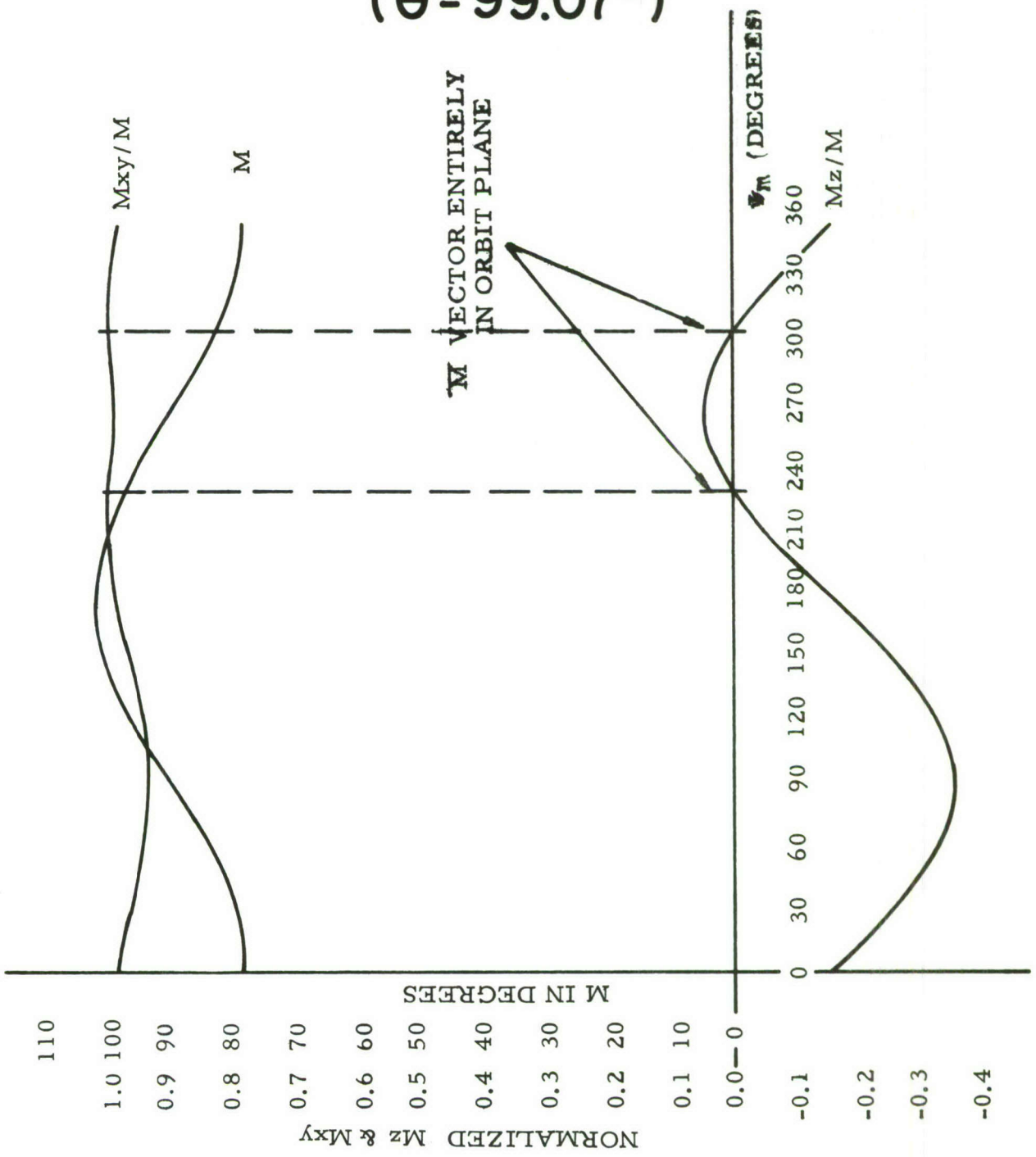
GRAPH 2.1

PLANER $\bar{B}$ FIELD PHASE AND MAGNITUDE VARIATION WITH ORBITAL POSITION (ϕ)

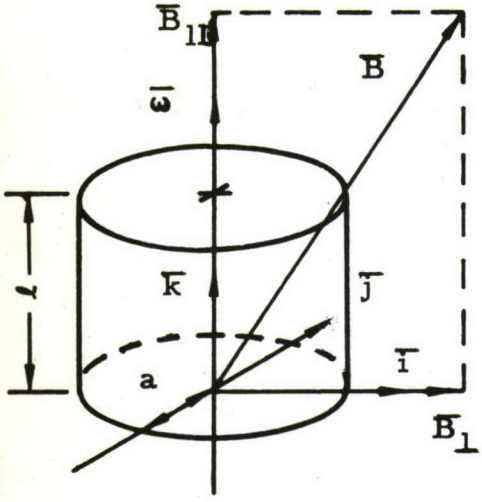


GRAPH 2.2

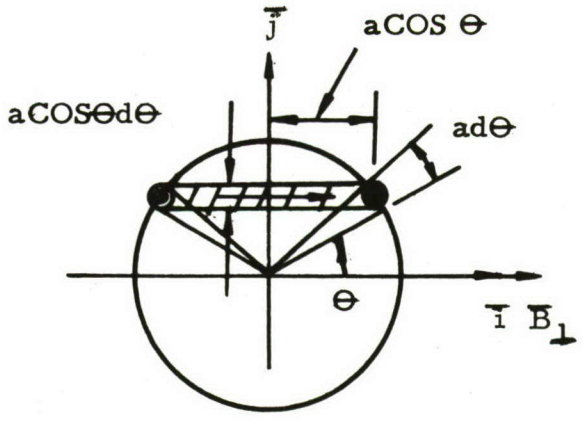
MAGNITUDE & PHASE OF NORMAL & PLANER DIPOLES VS ψ_m ($\theta = 99.07°$)



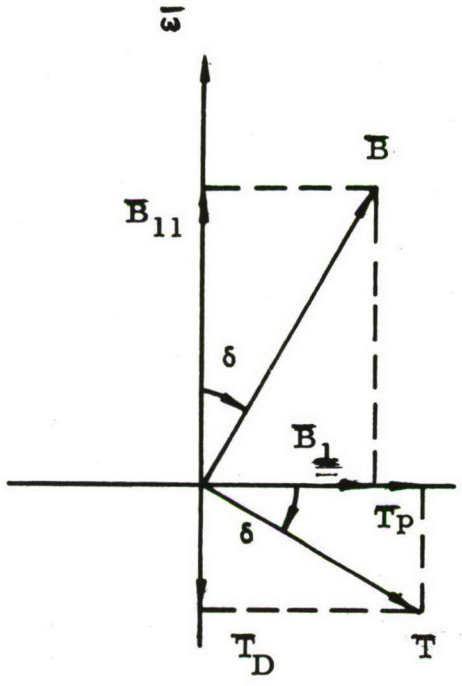
GRAPH 2.3


$\bar{B}_{11}$
$\bar{B}$
$\bar{\omega}$
$\bar{k}$
$\bar{j}$
$\bar{i}$
a
$\bar{B}_{\perp}$

(A)

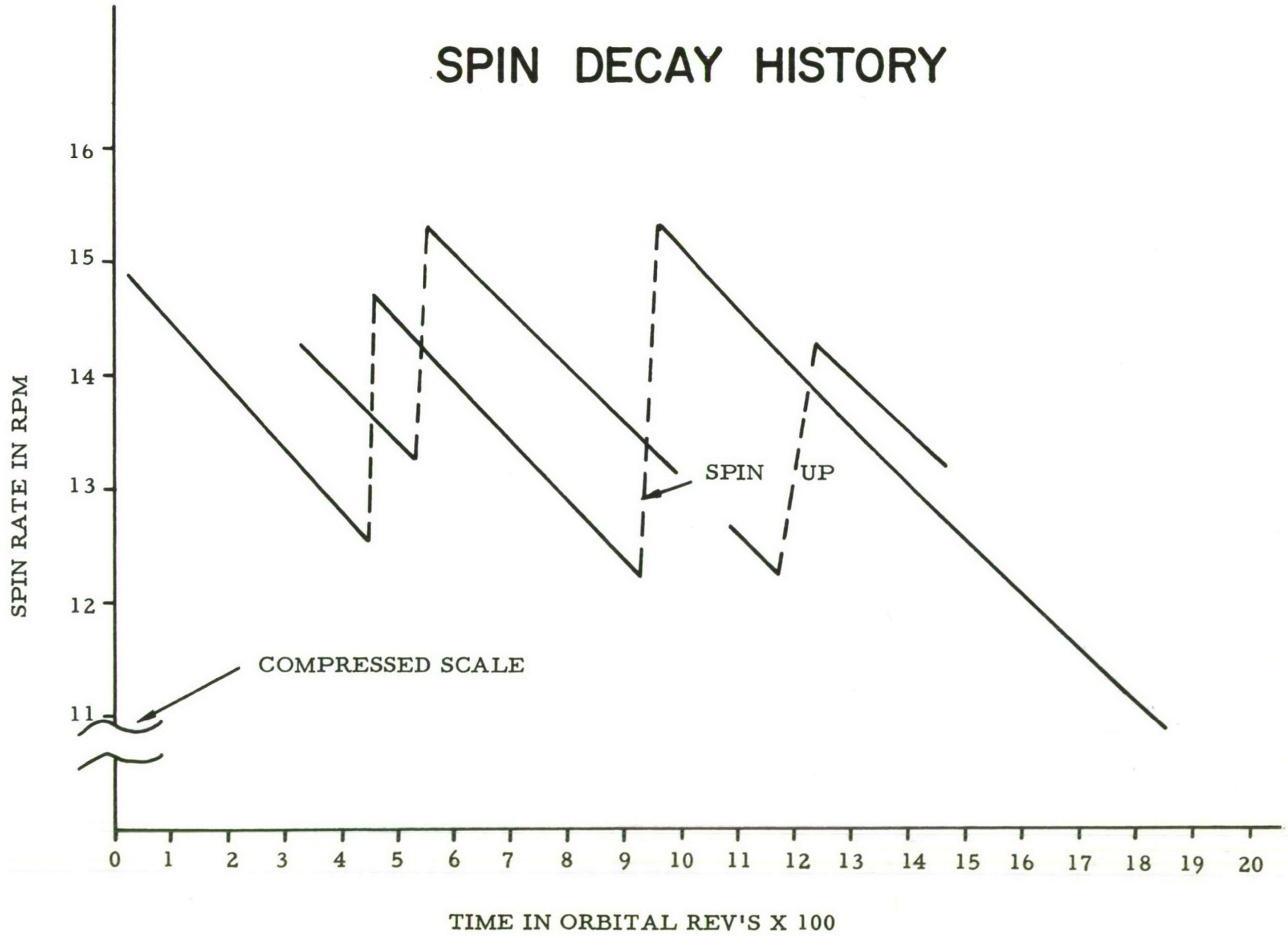

$\bar{j}$
aCOS θ
aCOSθdθ
adθ
θ
$\bar{i}$ $\bar{B}_{\perp}$

(B)


$\bar{\omega}$
$\bar{B}$
$\bar{B}_{11}$
δ
$\bar{B}_{\perp}$
T_P
δ
T_D
T

(C)

FIG. 3.1


SPIN DECAY HISTORY
SPIN RATE IN RPM
16
15
14
13
12
11
SPIN UP
COMPRESSED SCALE
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
TIME IN ORBITAL REV'S X 100

GRAPH 3.1

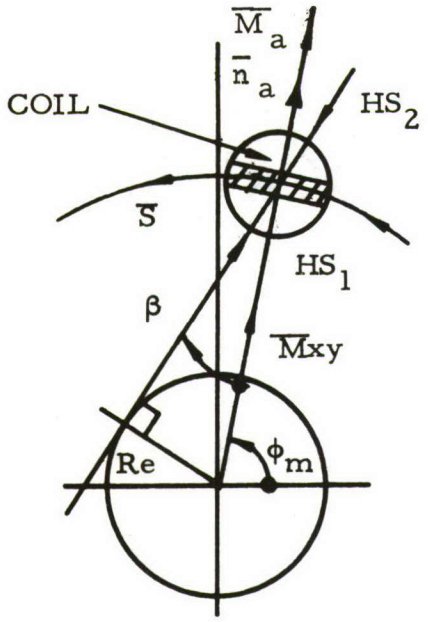


FIG. 5.1

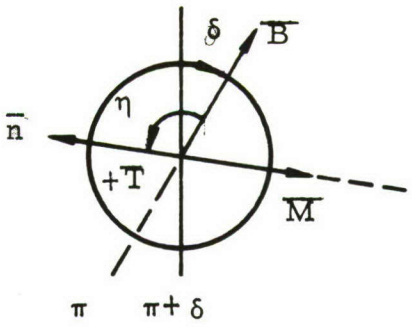


FIG. 5.3

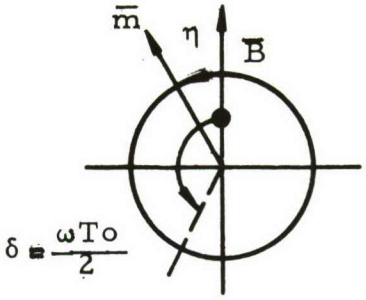
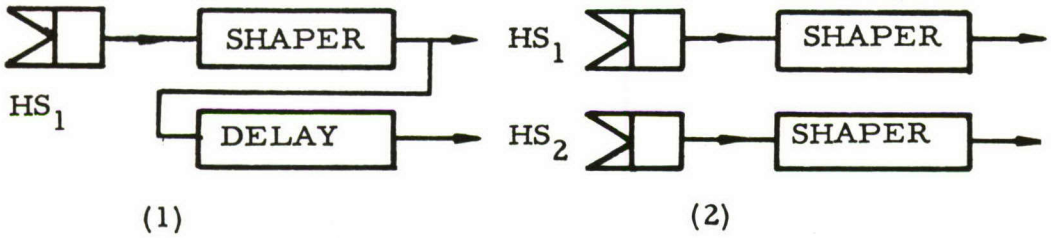


FIG. 5.4



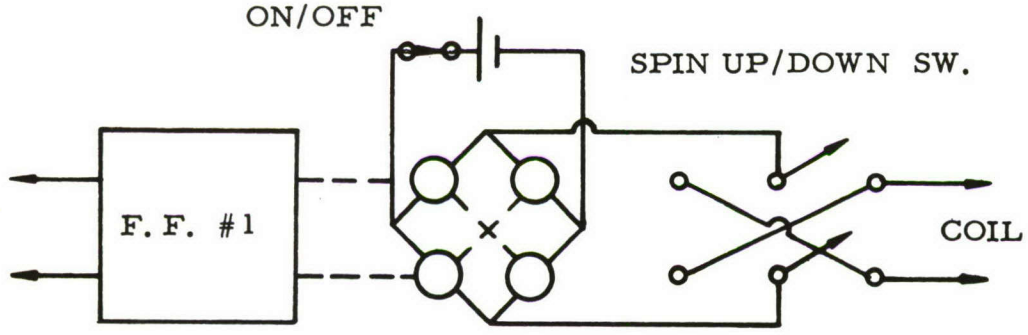


FIG. 5.2

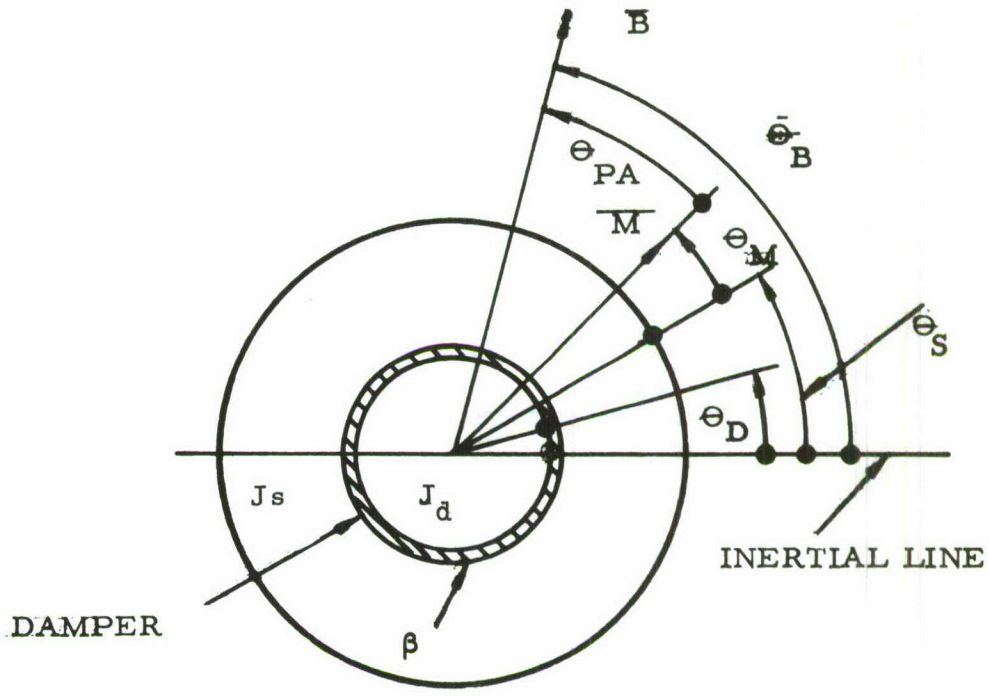
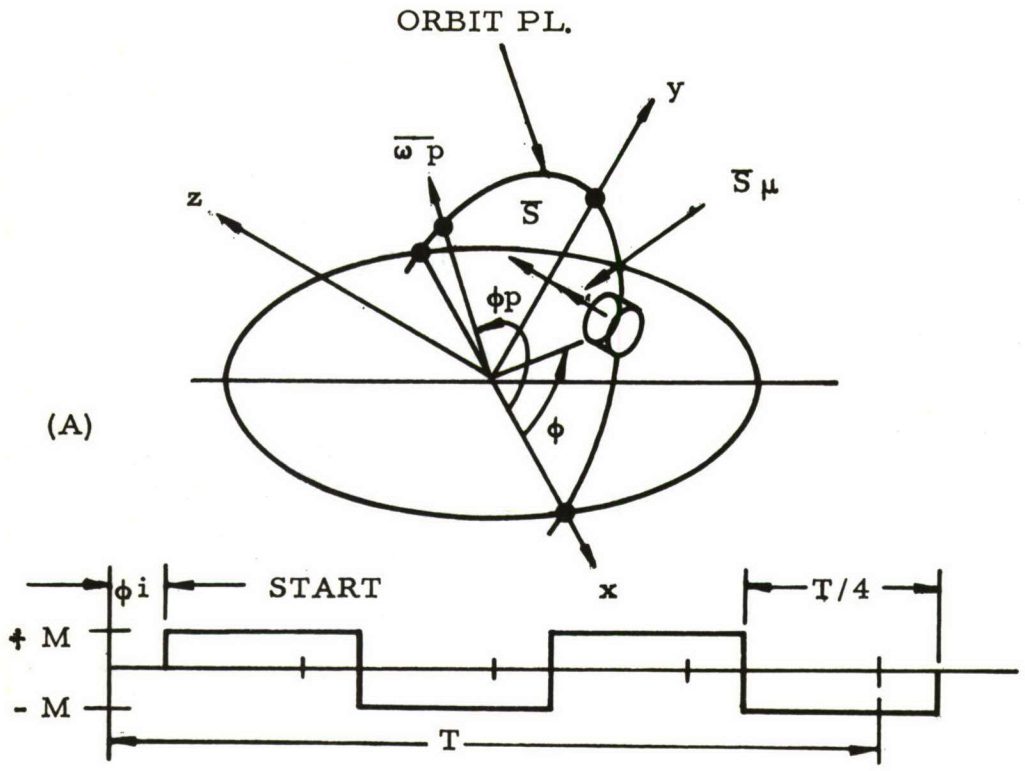


FIG. 6.1

TABLE 6.1

J_d	Damper Moment of Inertia (Kg - Meters^2)
Js	Satellite Moment of Inertia (Kg - Meters^2)
β	Viscous Drag Coefficient (Nt - Meters/Rad/Sec)
$\overline{B}$	Flux Density Vector (Webers/Meter^2)
$\overline{M}$	Dipole Moment Vector (Amp-Meter^2)



(B) COIL DIPOLE MOMENT

FIG. 7.1

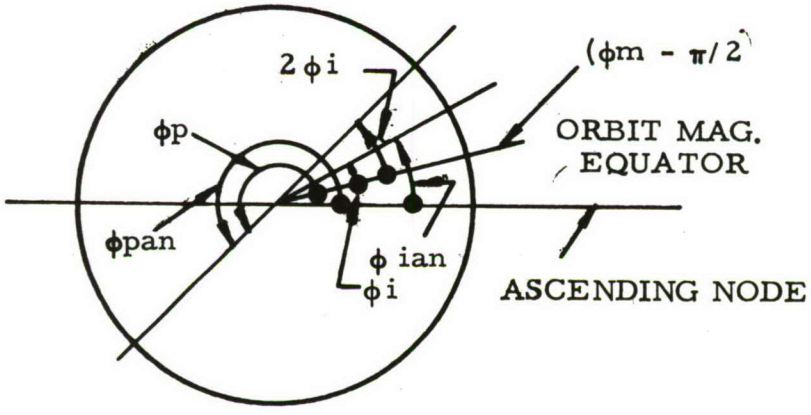


FIG. 7.2

TABLE 8.1

θ_a	= Angular position of motor amature
θ_s	= Angular position of satellite
θ_r	= Independent "Requirement" variable
θ_f	= Angular position of rotating motor field with respect to the motor (or satellite) frame.
$\epsilon=\theta_r-\theta_s$	= Error between instantaneous satellite position and desired position.
β	= Motor damping constant
J_a	= Armature moment of inertia
J_s	= Satellite moment of inertia
k_t	= Motor torque constant
k_p	= Proportional control constant
k_i	= Integral control constant.

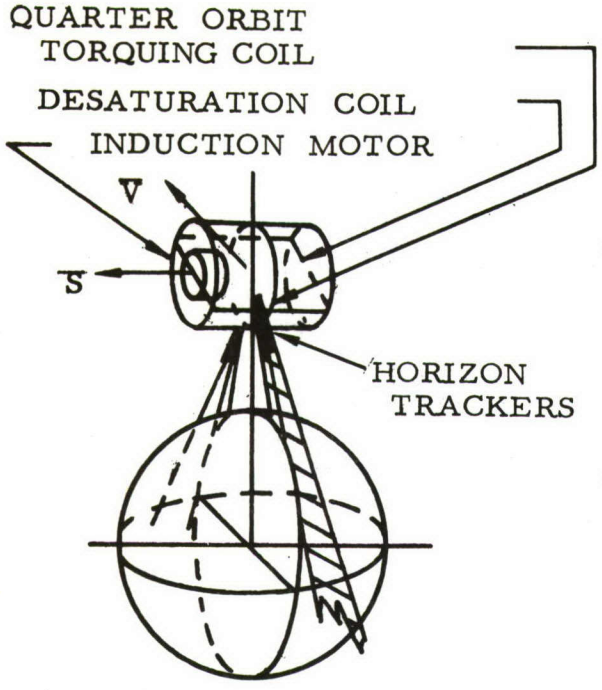


FIG. 8.1

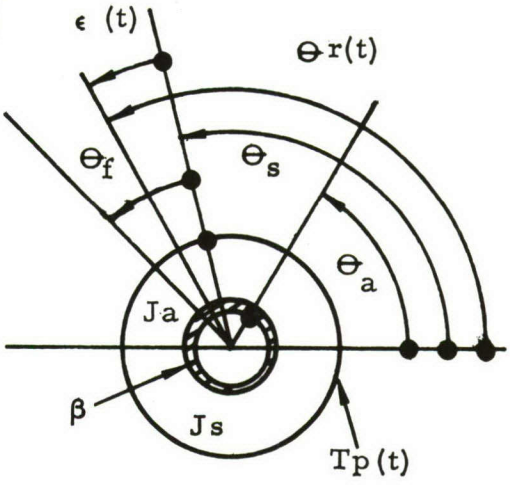


FIG. 8.3

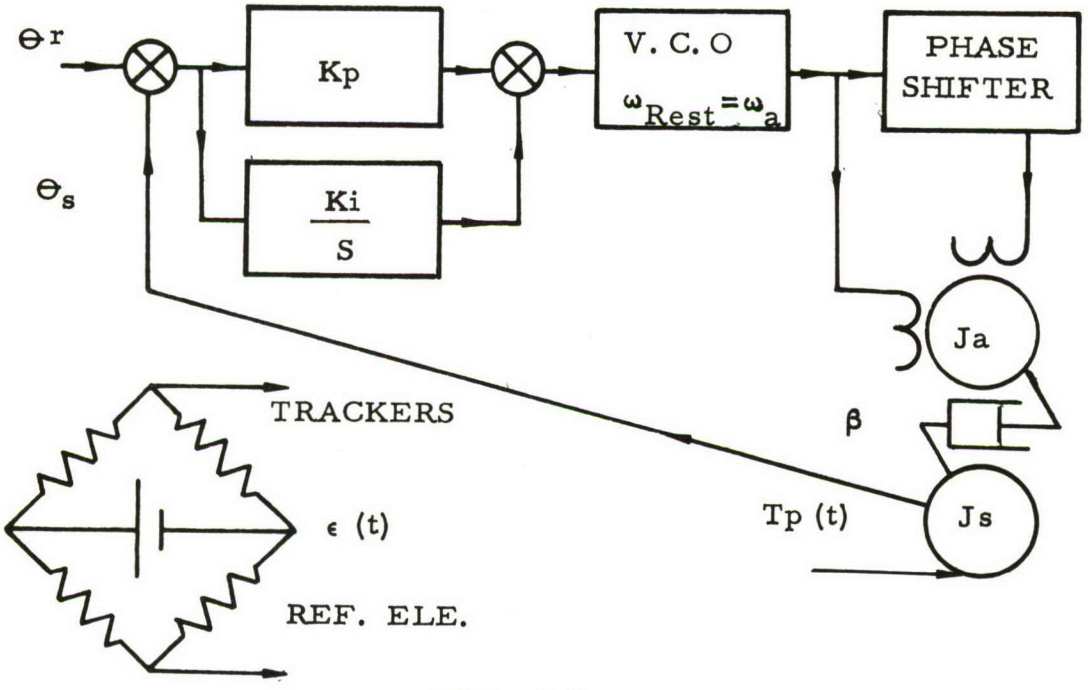


FIG. 8.2

BIBLIOGRAPHY

2. The Earth's Magnetic Field

1. Mlodnosky, R.F., and Helliwell, R.A., "Graphic Data on the Earth's Main Magnetic Field in Space", Journal of Geophysical Research, Vol. 67, No. 6, June 1962, pg. 2207.
2. U. S. Air Force, Air Research and Development Command, Handbook of Geophysics, Macmillan Co., N.Y., 1960, Chpt. 10.
3. Leyhe, Edward W., "Geomagnetic and Interplanetary Magnetic Field Environment of an Earth Satellite", NASA TN D-1019, July 1962, pg. 50.
4. Nelson, W.C., and Loft, E.E., Space Mechanics, Prentice-Hall Inc., Englewood Cliffs, N.J., 1962, pg 96.
5. Thomson, W. T., Introduction to Space Dynamics, John Wiley & Sons Inc., N.Y., 1961, pg 94 (After Blitzer)

3. Eddy Current Damping and Drag

1. Vinti, John P., "Theory of the Spin of a Conducting Satellite In the Magnetic Field of the Earth", Rep. No. 1020, Ballistic Res. Labs., Aberdeen Proving Ground, July 1957. (Also DDC No. AD 149,360)
2. Rosenstock, Herbert B., "The Effects of the Earth's Magnetic Field on the Spin of the Satellite", Astronautical Acta, Vol. III, Fasc. 3, 1957, pg 215.
3. Hooper, John W., "The Damping of Metalic Cylindrical and Sperical Bodies Rotating in a Uniform Magnetic Field", Tech. Rep. DG-R-15, Dev. Operations Div., Army Ballistic Missile Agency, Huntsville, Ala., Nov. 1957, (Also DDC No. AD 147,156)

4. Zonov, Yu. V., "On the Problem of the Interaction Between a Satellite and the Earth's Magnetic Field", NASA Tech. Translation F-37, translated from Iskusstvennyye Sputniki Zenili, No. 3, Academy of Science USSR, Moscow, 1959.

5. Wilson, Raymond H., Jr., "Rotational Magnetodynamics and Steering of Space Vehicles", XIth International Astronautical Congress, Stockholm, 1960, pg 525.

6. Wilson, R.H., "Magnetic Damping of Rotation of Satellite 1958 B2", Science, Sept 25, 1959, pg 791.

7. Wilson, R.H., "Geomagnetic Rotational Retardation of Satellite 1959 A1 (Vanguard II)", Science, Jan 22, 1960, pg 223.

8. LaPaz, Lincoln, "Magnetic Damping of Rotation of Vanguard I Satellite", Science, Feb 5, 1960, pg 355. (Followed by Wilson's reply to LaPaz's remarks)

9. Smythe, William R., Static and Dynamic Electricity, McGraw-Hill Book Co., New York, 1950, pg 417, Prob. 16.

10. Eide, Donald G., "Experimental Evaluation of the Deceleration of Aluminum Cylinders Rotating in a Magnetic Field and Comments on Magnetic Damping of a Flywheel Control", NASA Tech. Note D-749.

11. Grasshoff, L.H., "Eddy Current Torque Compensation in a Spin Stabilized Earth Satellite", ARS Journal, March 1961, pg. 290.

12. Williams, Donald D., "Torques and Attitude Sensing in Spin Stabilized Satellites", American Astronautical Society, Goddard Memorial Symposium, Mar 16, 1962, Washington D.C., AAS 62-43.

4. Hystersis Damping and Residual Dipole Moments

1. Fischell, Robert E., "Magnetic Damping of the Angular Motions of Earth Satellites", ARS Journal, Sep 1961, pg 1210.
2. Fischell, Robert E., "Magnetic and Gravity Attitude Stabilization of Earth Satellites", John Hopkins University, Applied Physics Laboratory, May 1961, CM-996 (Also DDC No. AD 271 383).

6. Synchronous Motor Satellite

1. Freeman, David J., "Automatic Latitude/Longitude Sensing Techniques for Earth Satellites", Aerospace Engineering, Nov 1962, pg 48.
2. WDL Div., Philco Corp., "Implementation Study for a Passive Damper", WDL Technical Proposal 1241 (WDL-TP1241), Palo Alto, Calif., 31 Jan 64.
3. Marx, S. H., Whitmore, A., and Blount, G., "Development of a Lossy Coupling Between Hinged Inertia Members for a Gravity Gradient Stabilized Spacecraft", WDL-TR 2212, Philco Corp., WDL Div., Palo Alto, Calif., 7 Jan 1964.

7. Two Degree Attitude Control (Spin Stabilized)

1. Colombo, C., "On the Motion of Explorer XI around its Center of Mass", Amer. Astn. Soc., Goddard Memorial Symposium, Mar 16, 1962, Washington D.C., 62-45.
2. Hecht, E., and Manger, W.P., "Magnetic Attitude Control of the Tiros Satellite", Amer. Astn. Soc., Goddard Memorial Symposium, Mar 16, 1962, 62-44.

3. Hertz, H., "Induction in Rotating Spheres", Miscellaneous Papers, Macmillan, London (1896), pg 155.

4. Vinti, J., "Theory of the Spin of a Conducting Satellite", Ballistic Research Laboratories Reports Nos. 1020 and 1031, 1957.

5. Smythe, W.R., Static and Dynamic Electricity, McGray Hill Book Co., Inc., New York, 1950, pg 307.

6. Grasshoff, L. H., "A Method for Controlling the Attitude of a Spin Stabilized Satellite", ARS Preprint 1501-60, 1960.

7. Grasshoff, L. H., "Eddy Current Torque Compensation in a Spin Stabilized Earth Satellite", ARS Journal, Mar 1961, pg 290.

8. Blakemore, D. J., "Magnetic Torquing Scheme", Notes Sect., AIAA Jour. Vol. I, No. 8, pg 1888.

9. Fischell, Robert E., "Magnetic and Gravity Attitude Stabilization of Earth Satellites", John Hopkins University, Applied Physics Laboratory, May 1961, CM 996 (Also DDC No. AD 271 383).

10. Kamm, L. J., "Magnetorquer--A Satellite Orientation Device", ARS Journal, June 1961, pg 813.

11. White, John S., Shigemoto, Fred H., and Bowiquin, K., "Satellite Attitude Control Utilizing the Earth's Magnetic Field", NASA Technical Note D-1068.

12. Buckingham, A.G., "A New Method of Attitude Control Utilizing the Earth's Magnetic Field for Long Life Space Vehicles", Guidance, Control and Navigation Conference, Stanford University, Stanford, Calif., Aug 7-9, 1961, ARS Reprint 1915-61.

8. Three Degree Attitude Control (One Storage Element)

1. Harnwell, Gayload P., Principles of Electricity and Electromagnetism, McGraw-Hill Book Co. Inc., N.Y., 1949, pg 450.
2. Gardner, M.F. and Barnes, J.L., Transients in Linear Systems, Vol. I, John Wilesy & Sons, Inc., N.Y., 1942, pg 343. Transform 1.310.
3. Cannon, Robert H., "Some Basic Response Relations for Reaction-Wheel Attitude Control" ARS Journal, Jan. 1962, pg 61.
4. DeBra, D.B., and Cannon, R.H., "Momentum Vector Considerations in Wheel-Jet Satellite Control System Design", ARS Guidance, Control and Navigation Conference, Stanford University, Stanford, Calif., Aug. 7-9, 1961, ARS 1917-61.
5. Berstein, T., and Howard, D.R., "Design Considerations for Space Vehicle Reaction Control A-C Servomotors", ARS Guidance, Control, and Navigation Conference, Stanford University, Stanford, Calif Aug. 7-9, 1961, ARS 1916-61.
6. Vaeth, James E., "Flywheel Control of Space Vehicles", IRE International Convention Record, Part 4, IRE International Convention, N.Y., Mar 21-24, 1960, pg 91.
7. Ormsby, R.D., and Smith, M.C., "Capabilities and Limitations of Reaction Spheres for Attitude Control", ARS Journal, Jun 1961, pg 808.
8. Hering, K.W., and Hufnagel, R.E., "Inertia Sphere System for Complete Attitude Control of Earth Satellites", ARS Journal, Aug 1961, pg 1074.

APPENDIX A

ATTITUDE ERROR DETECTION SYSTEM*

The position of the Spin Vector in the orbit reference frame (xyz) is determined by two infrared horizon sensors mounted on the satellite as shown in Figure A.1. As the satellite spins, the cones generated by the horizon sensors intersect the earth and generate "earth contact pulses." The instantaneous roll error is computed from the duration of these pulses. Because roll and yaw error are cyclically interchanged during an orbit, the instantaneous roll error is insufficient to determine the inertial position of the spin vector. However, through the use of two values of the instantaneous roll error, the orientation of the spin vector may be computed. Knowing the orientation of the spin vector then permits the proper phasing of the quarter orbit torquing current.

The derivation of the position of the spin vector utilizes the geometry shown in Fig. A.2. The orbit plane is the (xy) plane in the figure. The normal to the plane (determined in a right hand sense) is z. The primed coordinates are based upon the satellite velocity vector, orbit normal, and a ray to the geocenter (this implicitly a circular orbit). The yaw error is ψ and the roll error is φ. The vector $\overline{S}$ (the spin vector) is assumed to be normalized. Let the primed and unprimed frame be related by the matrix (A_t).

* The analysis reported here was performed in part by J. A. Mallin under Air Force Contract No. AF04(695)-15.

From Fig. A.2 the x and y direction cosines in the primed coordinates are:

$$S_x' = \mathrm{Sin}\varphi$$
$$S_y' = \mathrm{Sin}\psi \qquad (A.1)$$

In the xyz frame, the unit spin vector is:

$$\bar{S} = S_x\bar{i}+S_y\bar{j}+S_z\bar{k} \qquad (A.2)$$

The sine of the roll error in unprimed coordinates which is a constant and the maximum value obtained by φ in satellite coordinates, is:

$$\mathrm{Sin}\varphi_{max} = \sqrt{S_x^2+S_y^2} \qquad (A.3)$$

Equating the components of $\bar{S}$ in the two coordinate systems to one another with the aid of (A_t) and applying the results of Eq.s (A.1) and A.3):

$$\mathrm{Sin}\varphi = \mathrm{Sin}\varphi_m\mathrm{Sin}(\theta+\lambda)$$
$$\mathrm{Sin}\psi = \mathrm{Sin}\varphi_m\mathrm{Cos}(\theta+\lambda) \qquad (A.4)$$

Where $\lambda = \mathrm{Tan}^{-1} S_x/S_y$ and is called the "orbital phasing" of the satellite attitude error. If φ_m is small, then so too is ψ and φ so that Eq. (A.4) becomes:

$$\varphi = \varphi_m\mathrm{Sin}(\theta+\lambda)$$
$$\psi = \varphi_m\mathrm{Cos}(\theta+\lambda) \qquad (A.5)$$

If two different values of φ are known corresponding to two different orbital positions θ, then φ_m and λ can be computed. The value of λ determines the phasing of the quarter orbit torquing current and φ_m is

the arc through which the spin vector is to be precessed. The value of θ when φ is equal to φ_m is:

$$\begin{aligned} \theta &= (\pi/2 - \lambda) \\ \varphi &= \varphi_m \end{aligned} \tag{A.6}$$

ψ is zero at this time indicating that the attitude error appears entirely as a roll error.

The instantaneous roll error φ is determined from the duration of earth contact pulses by use of the formula:

$$\varphi = \mathrm{Tan}^{-1}\left[\frac{\mathrm{Tan}\alpha[\mathrm{Cos}\beta - \mathrm{Cos}\beta']}{2}\right] \tag{A.7}$$

β and β' are obtained from the horizon sensor data as depicted in Figure (A.3).

T_e is the earth contact period and T_s is the spin period of the satellite.

Eq. (A.7) is derived with the aid of Fig. A.4. The various symbols in the figure are defined as:

h = orbit altitude

$\mu = \mathrm{Sin}^{-1}(1+h/R_e)^{-1}$

α = angle between spin axis and horizon sensor

β = one half earth contact angle

φ = instantaneous roll error

R_e = radius of earth

$\bar{I}$ = vector from satellite to point of tangency to the earth. (Where the horizon sensor ray commences to enter or leave the earth).

The W axis component of the $\bar{I}$ vector is computed in two different ways. The first way is simply $I_W = I\text{Cos}\mu$. The second method involves projecting $\bar{I}$ on the various construction lines shown in the figure. Since the W components computed by the two methods must be identical, equating them and simplifing the results yields:

$$\text{Cos}\beta = \frac{\text{Cos}\mu}{\text{Sin}\alpha\text{Cos}\varphi} + \frac{\text{Tan}\varphi}{\text{Tan}\alpha} \qquad \text{(A.8)}$$

Eq. (A.8) is the "single scanner" case. Since μ and α are known by the orbit and satellite geometry, φ may be determined given β from the telemetry data.

With two scanners, placed as shown in Figure (A.1), the roll angle for the second scanner is the negative of the first. Replacing φ by $-\varphi$ in Eq. (A.8) then yields the corresponding equation for the second scanner.

$$\text{Cos}\beta' = \frac{\text{Cos}\mu}{\text{Sin}\alpha\text{Cos}\varphi} - \frac{\text{Tan}\varphi}{\text{Tan}\alpha} \qquad \text{(A.9)}$$

Subtracting Eq. (A.9) from Eq. (A.8) yields the "two scanner" equation.

$$\text{Cos}\beta - \text{Cos}\beta' = \frac{2\ \text{Tan}\varphi}{\text{Tan}\alpha} \qquad \text{(A.10)}$$

which is an equivalent form of Eq. (A.7).

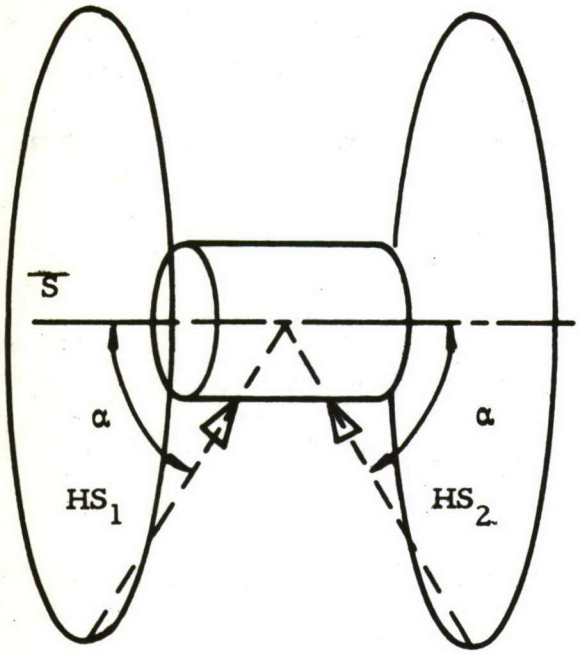


FIG. A.1

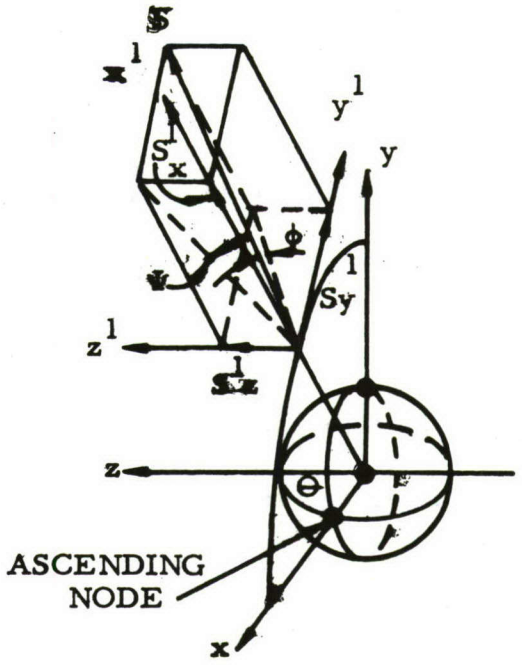


FIG. A.2

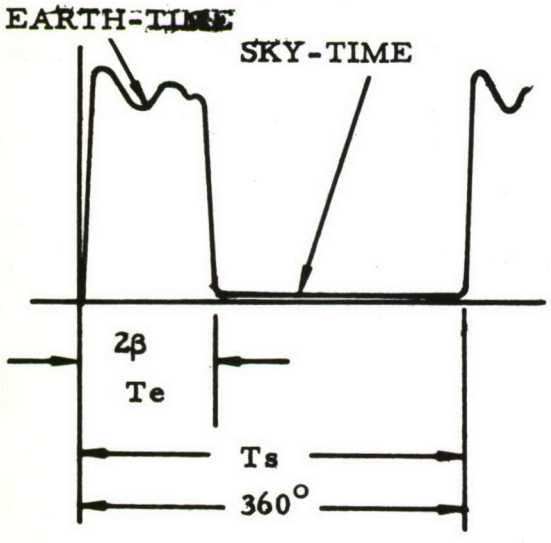


FIG. A.3

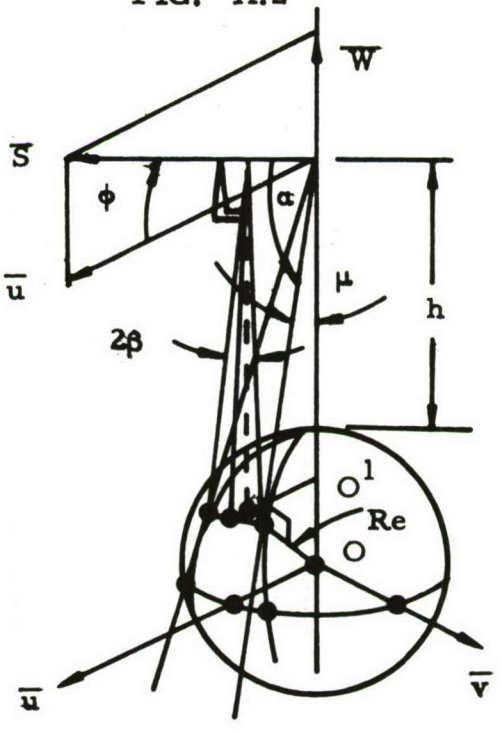


FIG. A.4

APPENDIX B

COIL DESIGN CONSIDERATIONS

The magnetic dipole moment of an (N) turn coil of area (A) carrying a current (I) is:

$$M = NIA \tag{B.1}$$

In practice, M is a function of the coil weight, the power dissipated in the resistance of the coil, and the physical properties of the conductor. A "payoff" equation involving these factors will be derived.

Let r_c = coil radius

d_ω = wire diameter

ρ = resistivity of conductor material

μ = density of conductor material

From the calculus of variation it is known that the circle circumscribes the greatest area for a given perimeter. Therefore we assume the form of the coil to be a circle.

The resistance of the coil is:

$$R_c = \frac{8Nr_c\rho}{d\omega^2} \tag{B.2}$$

The power dissipated in the coil is:

$$P_c = I^2R_c = \frac{8N_cr_c\rho I^2}{d\omega^2} \tag{B.3}$$

The mass of the conductor is:

$$M_c = \frac{Nr_c\mu d\omega^2\pi^2}{2} \tag{B.4}$$

Substituting (B.3) and (B.4) into Eq. (B.1) and solving for M, the diople moment:

$$M = \frac{r_c p_c^{\frac{1}{2}} M_c^{\frac{1}{2}}}{2(\mu\rho)^{\frac{1}{2}}} \qquad (B.5)$$

For a given mass and power dissipation, Eq. (B.5) emphasizes the desirability of maximizing r_c and minimizing the $\mu\rho$ product of the conductor. The $(\mu\rho)$ products of the common conductors are listed in Table B.1. They are normalized with respect to copper. Since aluminun has the lowest value of $\mu\rho$ product, aluminun wire is chosen for the coil.

TABLE B.1

Material	$(\mu\rho)$ Product
Copper	1.0
Silver	1.12
Aluminun	0.50
Magnesium	0.53

A representative coil might consist of 780 turns of #32 aluminum wire. The coil radius is 11.7 inches, its resistance is 1800 ohms and the current is 6 milliamperes. This results in a 1.75 amp-meter^2 dipole moment. The power consumed is 65 milliwatts.

APPENDIX C

EDDY CURRENT DAMPER COIL DESIGN

Several authors have suggested the use of an (N) turn coil with shorted terminals to provide eddy current damping of a satellite. The resulting, energy dissipation equation is of the general form:

$$\frac{d(\text{K.E. RMS})}{dt} = -\frac{N^2A^2B^2\omega^2}{2}\,\frac{R}{(R^2+\omega^2L^2)} \qquad \text{(C.1)}$$

Where R = coil resistance

W = spin rate

L = coil inductance

N = no. of turns

A = area of loop

B = flux density

Though Eq. (C.1) is "correct", it is also very misleading. Since performance per unit weight must be maximinized in satellite design, the question is : "given x pounds of conductor, how should this be extruded for maximum damping?" A single turn is the optimum geometry. Were it not for practical considerations (N) turns and 1 turn would give identical results. However, in the (N) turn coil the weight and volume of the insulation, as well as the weight of the necessarily more protective coil support are completely parasitic.

To illustrate the theoretical equality of the 1 turn and (N) turn case we note that the inductance of a coil varies as the square of its

turns (Smythe pg 318). The resistance of a given volume $V=l_c A_c$ of material is:

$$R = \frac{\rho l_c}{A_c} \tag{C.2}$$

If the same amount of material is extruded to a new length $l_c'=Nl_c$ and reduced crossection $A_c'=\frac{A_c}{n}$ the resistance is:

$$R' = \frac{\rho l_c'}{A_c'} = \frac{\rho(Nl_c)}{(A_c/N)} = \frac{N^2\rho l_c}{A_c} = N^2 R \tag{C.3}$$

Substitution of these results for R and L in Eq. (C.1) demonstrates the theoretical equivalence of coils of any number of turns.

At low spin rates where ωL is much less than R, Eq. (C.1) simplies to:

$$\frac{d\ (\text{K.E. RMS})}{dt} = \frac{A^2B^2\omega^2}{2R} \tag{C.4}$$

A tradeoff equation can be derived from Eq. (C.4) by introducing the relations:

$$M_{(\text{mass})} = \mu V(\text{Volume}) = \mu l_c A_c \tag{C.5}$$

into Eq. (C.4) obtaining:

$$\frac{d(\text{K.E. RMS})}{dt} = \frac{r^2 M B^2 \omega^2}{8\mu\rho} \tag{C.6}$$

Where r is the radius for a circular coil. Eq. C.6 indicates that for a given mass of conductor, r should be made as large as possible. Again, as in the case for the attitude control coil, the $\mu\rho$ product should be minimized suggesting the use of aluminum.

A Method for Controlling the Attitude of a Satellite Using the Earth's Magnetic Field

Wilfred Joseph Minkus

GS-12

Applied Research Section (EMCRR)
Communications Research Branch
Communications Division
Rome Air Development Center
Griffiss Air Force Base, New York

TABLE OF CONTENTS

1.	Introduction	1
2.	Transformations and Coordinates	3
3.	Equations of Motion	6
4.	The Time-Optimum Control Equations	10
5.	The Control Required to Minimize Cross-Coupled Torques	17
6.	RADC Applications	25
7.	Concluding Remarks	30
	Appendix I	32
	Appendix II	36
	Glossary of Symbols	39

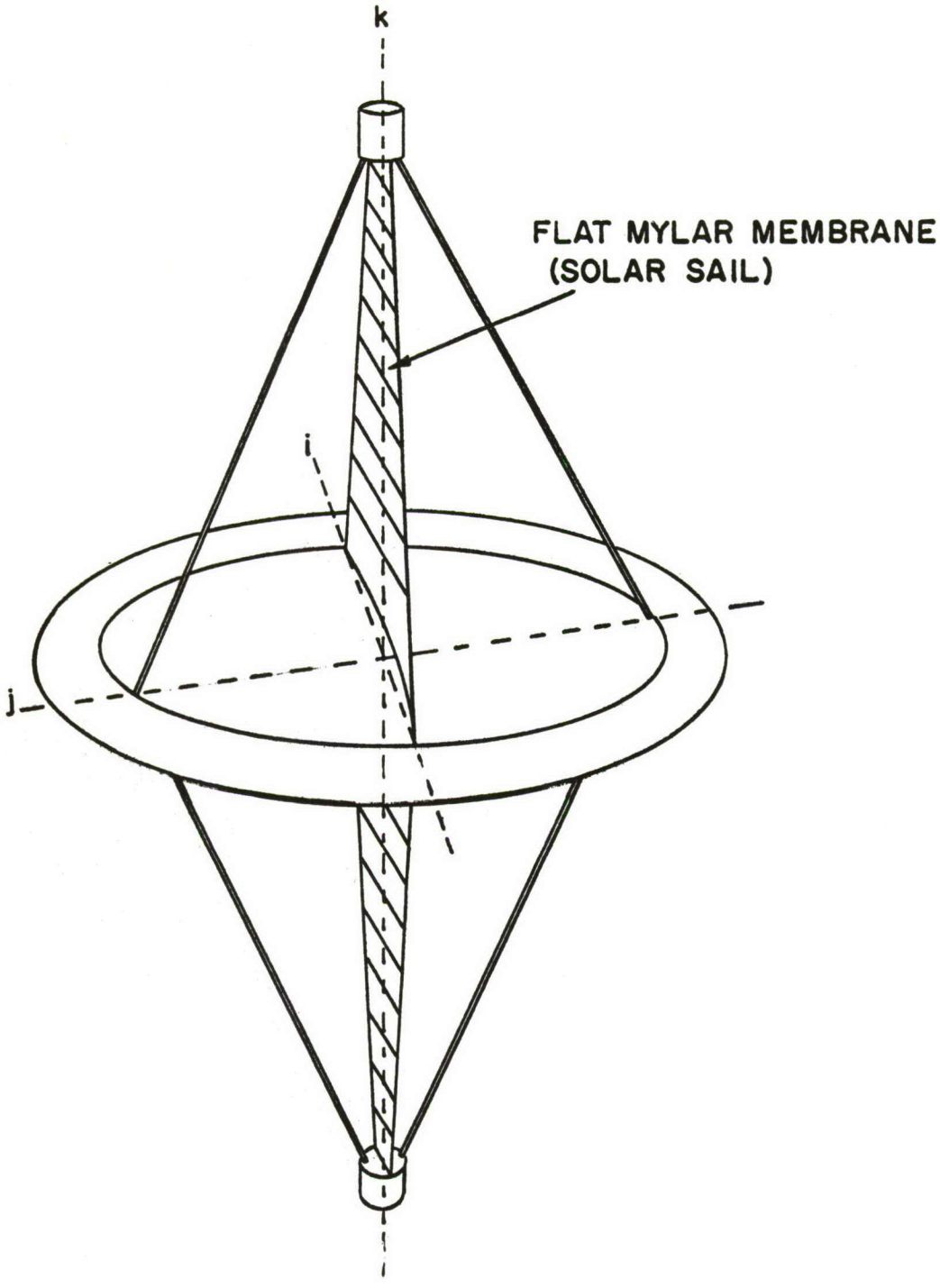
1. INTRODUCTION

This paper investigates a means of controlling the attitude of a satellite in a circular orbit about the earth by the torque due to the interaction between the earth's magnetic field and magnetic moments produced by currents in three orthogonal coils on board the satellite. This system has the advantage over more conventional ones in that no mass is expended in the control process. It is designed to rotate the satellite about its longitudinal axis in a minimum amount of time while limiting the deviation of this axis from the local normal to the earth.

The analysis is limited to the period during which the satellite's attitude is being changed by magnetically produced torques. When these torques are not applied, gravity gradient effects act as a restoring force to align the longitudinal axis with the local normal to the earth. It is assumed that the deviation and the rate of deviation of this axis from the earth's normal are negligibly small at the beginning of the control period.

The method is applicable to a satellite with symmetry such that the moment of inertia about its longitudinal axis is much smaller than those about body-fixed axes lying in a plane perpendicular to it and whose moments of inertia about axes in this plane are approximately equal. The model of the earth's magnetic field that is used is a dipole with its poles located at the observed geographical positions.

The Rome Air Development Center is currently interested in highly reliable random access world-wide military communication systems. This capacity could be provided by a system of satellites employing passive-type signal processing devices on board and using solar pressure to provide



SATELLITE SKETCH

FIGURE 1.

orbital control. The solar pressure would impinge on a flat membrane, a solar sail, whose attitude with respect to the sun line is set by the proposed control system. A type of solar sailing satellite, and one that is of particular interest to RADC, is shown in Figure 1.

The control system is developed to be applicable to the general type of satellite indicated above and to all orbits except those passing close to the magnetic poles. It is then applied to a particular satellite and type of orbit of interest to RADC.

2. TRANSFORMATIONS AND COORDINATES

The configuration of the satellite and the location of its principal axes, designated i, j, and k, are indicated in Figure 1. It will be noted that these axes form a right-handed triad with the k-axis taken as the longitudinal axis and their origin at the center of gravity of the satellite. This set will be referred to as the body-fixed reference frame. Another right-handed triad is referenced to this center of gravity and designated as I, J and K. This system will be referred to as the orbit reference frame. The I-axis points along the orbit in the direction of motion, the J-axis is normal to the plane of the orbit, and the K-axis points away from the earth along its local normal. For the spherical model of the earth assumed in this analysis, this normal is a radial line from the center. Referring to the angles designated in Figure 2, an Euler transformation from the orbital to the body fixed reference frame is:

$$(1) \qquad E = \begin{bmatrix} \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ (\sin\psi\sin\theta\cos\phi) - \cos\psi\sin\phi & (\sin\psi\sin\theta\sin\phi) + \cos\psi\cos\phi & \sin\psi\cos\theta \\ (\cos\psi\sin\theta\cos\phi) + \sin\psi\sin\phi & (\cos\psi\sin\theta\sin\phi) - \sin\psi\cos\phi & \cos\psi\cos\theta \end{bmatrix}$$

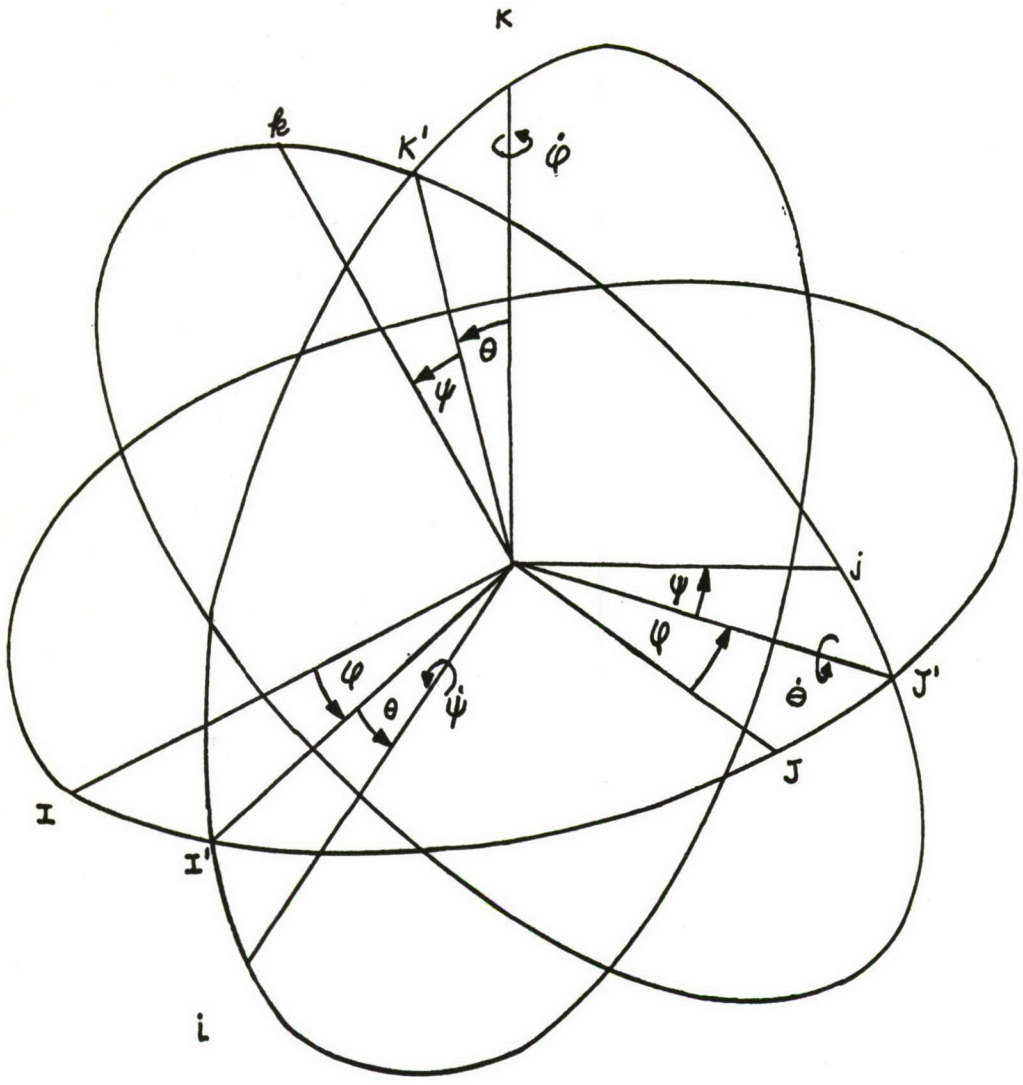


FIGURE 2. EULER TRANSFORMATION COORDINATES AND ANGLES

The transformation from $\dot{\theta}, \dot{\phi}$ and $\dot{\psi}$ to the angular velocity about the principal body-fixed axis ω_i, ω_j and ω_k is:

$$(2) \qquad A = \begin{bmatrix} 1 & 0 & -\sin\theta \\ 0 & \cos\psi & \sin\psi\cos\theta \\ 0 & -\sin\psi & \cos\psi\cos\theta \end{bmatrix}$$

The angular velocity of the orbital reference frame with respect to one fixed at the earth is ω_o and is, of course, directed along the J-axis. Adding this effect to that produced by changes in the Euler transformation angles yields:

$$(3) \qquad \begin{bmatrix} \omega_i \\ \omega_j \\ \omega_k \end{bmatrix} = A \begin{bmatrix} \dot{\psi} \\ \dot{\theta} \\ \dot{\phi} \end{bmatrix} + E \begin{bmatrix} 0 \\ \omega_o \\ 0 \end{bmatrix}$$

Since the orbit is circular and the earth is assumed spherical, $\dot{\omega}_o = 0$,

$$(4) \qquad \begin{bmatrix} \dot{\omega}_i \\ \dot{\omega}_j \\ \dot{\omega}_k \end{bmatrix} = \dot{A} \begin{bmatrix} \dot{\psi} \\ \dot{\theta} \\ \dot{\phi} \end{bmatrix} + A \begin{bmatrix} \ddot{\psi} \\ \ddot{\theta} \\ \ddot{\phi} \end{bmatrix} + \dot{E} \begin{bmatrix} 0 \\ \omega_o \\ 0 \end{bmatrix}$$

For $|\theta| << 1$, $|\psi| << 1$ these relations become:

$$(5) \qquad \begin{bmatrix} \omega_i \\ \omega_j \\ \omega_k \end{bmatrix} = \begin{bmatrix} \dot{\psi} - \theta\dot{\phi} + \omega_o \sin\phi \\ \dot{\theta} + \psi\dot{\phi} + \omega_o \cos\phi \\ \dot{\phi} + \omega_o(\theta\sin\phi - \psi\cos\phi) \end{bmatrix}$$

$$(6)\quad \begin{bmatrix} \dot{\omega}_i \\ \dot{\omega}_j \\ \dot{\omega}_k \end{bmatrix} = \begin{bmatrix} \ddot{\psi} - \theta\,\ddot{\phi} - \dot{\theta}\,\dot{\phi} + \dot{\phi}\,\omega_o \cos\phi \\ \ddot{\theta} + \dot{\psi}\,\dot{\phi} + \dot{\psi}\,\dot{\phi} - \dot{\phi}\,\omega_o \sin\phi \\ \ddot{\phi} + \omega_o(\psi\dot{\phi} + \dot{\theta})\sin\phi + \omega_o(\theta\,\dot{\phi} - \dot{\psi})\cos\phi \end{bmatrix}$$

3. EQUATIONS OF MOTION

The equations of motion of the satellite about the body-fixed axes are:

$$(7)\quad \begin{bmatrix} T_i \\ T_j \\ T_k \end{bmatrix} = \begin{bmatrix} I_i\,\dot{\omega}_i + (I_k - I_j)\,\omega_k\,\omega_j \\ I_j\,\dot{\omega}_j + (I_i - I_k)\,\omega_i\,\omega_k \\ I_k\,\dot{\omega}_k + (I_j - I_i)\,\omega_j\,\omega_i \end{bmatrix}$$

where: T_i, T_j, T_k are the torques about body-fixed axes

I_i, I_j, I_k are the moments of inertia about the body-fixed axes.

For $I_i = I_j = I_o$, $\left|\frac{\dot{\phi}}{\omega_o}\right| >> \theta$, $\left|\frac{\dot{\phi}}{\omega_o}\right| >> \psi$:

$$(8)\quad \begin{bmatrix} T_i \\ T_j \\ T_k \end{bmatrix} \doteq \begin{bmatrix} I_o(\ddot{\psi} - 2\dot{\theta}\dot{\phi} - \psi\dot{\phi}^2 - \theta\ddot{\phi}) + I_k\,\dot{\phi}(\dot{\theta} + \psi\dot{\phi} + \omega_o\cos\phi) \\ I_o(\ddot{\theta} + 2\dot{\psi}\,\dot{\phi} - \theta\,\dot{\phi}^2 + \psi\,\ddot{\phi}) - I_k\dot{\phi}(\dot{\psi} - \theta\dot{\phi} + \omega_o\sin\phi) \\ I_k\,\ddot{\phi} \end{bmatrix},$$

which for small deviations and rates of deviation of the logitudinal axis (the k-axis) of the satellite from the earth's normal (the K-axis) becomes:

$$(9)\quad \begin{bmatrix} T_i \\ T_j \\ T_k \end{bmatrix} = \begin{bmatrix} I_o\,\ddot{\psi} + I_k\,\omega_o\,\dot{\phi}\cos\phi \\ I_o\,\ddot{\theta} - I_k\,\omega_o\,\dot{\phi}\sin\phi \\ I_k\,\ddot{\phi} \end{bmatrix}$$

The terms of this equation which are proportional to the second derivative of the Euler angles represent changes in angular momentum about body-fixed axes; those proportional to $\omega_o \dot{\phi}$ represent the cross-coupled torques due to orbital rotation.

In absence of external torques, the longitudinal axis of the satellite is restored to the earth's normal by the gravity gradient effect. With an effectively damped gravity gradient mechanism, the deviation and rate of deviation of the longitudinal axis of the satellite from the earth's normal can be assumed to be negligible at the beginning of the control period.

In body-fixed coordination the torque due to an undamped gravity gradient effect is:

$$(10)\quad \begin{bmatrix} T_{ui} \\ T_{uj} \\ T_{uk} \end{bmatrix} = -\frac{3G_o m_o}{2\gamma_o^3 r_o^3} \begin{bmatrix} (I_j - I_k)\cos^2\theta \sin^2\psi \\ (I_i - I_k)\cos\psi \sin 2\theta \\ (I_j - I_i)\sin\psi \sin 2\theta \end{bmatrix}$$

where: G_o is Newton's gravitational constant

m_o is the mass of the earth

r_o is the radius of the earth

γ_o is the radius of the orbit in earth radii

setting $K_o = \frac{3G_o m_o}{r_o^3} = 4.61 \times 10^{-6}$ [seconds^{-2}]

and for $I_i = I_j = I_o$, $\theta \ll 1$, $\psi \ll 1$ yields:

$$(11)\qquad \begin{bmatrix} T_{ui} \\ T_{uj} \\ T_{uk} \end{bmatrix} \doteq -K_o \gamma_o^{-3} (I_o - I_k) \begin{bmatrix} \psi \\ \theta \\ o \end{bmatrix}$$

Assuming the damping torque is proportional to angular velocity (viscous damping), the damped gravity gradient effect can be approximated as:

$$(12)\qquad \begin{bmatrix} T_{gi} \\ T_{gj} \\ T_{gk} \end{bmatrix} = -K_o \gamma_o^{-3} (I_o - I_k) \begin{bmatrix} \psi + H_o \dot{\psi} \\ \theta + H_o \dot{\theta} \\ o \end{bmatrix}$$

The same proportionality constant, H_o, has been used in the expressions for T_{gi} and T_{gj} with the effect of damping regarded as negligible about the longitudinal (or k-) axis.

The satellite is assumed to be equipped with a set of coils placed orthogonally to each other, each producing a component of magnetic moment along a body-fixed axis. The components of torque produced by the interaction of the components of magnetic moments and the earth's magnetic field about the body-fixed axes are:

$$(13)\qquad \begin{bmatrix} T_{Bi} \\ T_{Bj} \\ T_{Bk} \end{bmatrix} = \begin{bmatrix} o & B_k & -B_j \\ -B_k & o & B_i \\ B_j & -B_i & o \end{bmatrix} \begin{bmatrix} m_i \\ m_j \\ m_k \end{bmatrix}$$

where: m_i, m_j and m_j are the components of the magnetic moment

B_i, B_j and B_k are the components of the earth's magnetic field.

Since:

$$(14)\quad \begin{bmatrix} T_i \\ T_j \\ T_k \end{bmatrix} = \begin{bmatrix} T_{gi} \\ T_{gj} \\ T_{gk} \end{bmatrix} + \begin{bmatrix} T_{Bi} \\ T_{Bj} \\ T_{Bk} \end{bmatrix}$$

one may write using this and equations (9),(12) and (13):

$$(15)\quad I_o \ddot{\psi} + (K_o \gamma^{-3})(I_o - I_k)(\psi + H_o \dot{\psi}) = -I_k \dot{\phi} \omega_o \cos\phi + m_j B_k - m_k B_j$$

$$(16)\quad I_o \ddot{\theta} + (K_o \gamma^{-3})(I_o - I_k)(\theta + H_o \dot{\theta}) = I_k \dot{\phi} \omega_o \sin\phi + m_k B_i - m_i B_k$$

$$(17)\quad I_k \ddot{\phi} = m_i B_j - m_j B_1$$

The dynamics represented by equation (15), (16) and (17) can easily be written with reference to orbit coordinates under the assumption that the longitudinal axis is aligned close to the local normal to the earth with this assumption, the resultant of the components of torque represented by $I_o \omega_o \dot{\phi} \cos\varphi$ and $I_o \omega_o \dot{\phi} \sin\varphi$ is along the orbital direction, the I-axis, and equal to $I_o \omega_o \dot{\phi}$. The torque due to the interaction of magnetic moments of coils and magnetic field of the earth can be represented in orbit coordinates by replacing the components of these quantities in body-fixed coordinates by those in the orbit

frame. Finally, since it has been assumed that $I_i = I_j = I_o$ the moment of inertia about any axis in the i - j plane is equal to I_o. Using this fact and the alignment of the satellite's longitudinal axis along the earth's normal permits one to write that $I_I = I_J = I_o$ and $I_K = I_k$. Employing these results the dynamics represented by equations (15), (16) and (17) can be expressed as:

$$(18)\quad \ddot{\psi} + a_o H_o \dot{\psi} + a_o \psi = -\left(\frac{I_k}{I_o}\right)\dot{\phi}\omega_o + \left(\frac{1}{I_o}\right)\left(m_J B_K - m_K B_J\right)$$

$$(19)\quad \ddot{\theta} + a_o H_o \dot{\theta} + a_o \theta = \left(\frac{1}{I_o}\right)\left(m_K B_I - m_I B_K\right)$$

$$(20)\quad \dot{\phi} = \left(\frac{1}{I_k}\right)\left(m_I B_J - m_J B_I\right)$$

where

$$(21)\quad a_o = K_o \gamma^{-3}\left(1 - \frac{I_k}{I_o}\right) = 4.61 \times 10^{-6}\gamma^{-3}\left(1 - \frac{I_k}{I_o}\right)$$

4. THE TIME-OPTIMAL ATTITUDE CONTROL EQUATIONS:

For the condition regarding the alignment of the satellite's longitudinal axis, the relationship between the components of magnetic moments in orbital coordinates and those in body-fixed coordinates is:

$$(22)\quad \begin{bmatrix} m_I \\ m_J \\ m_K \end{bmatrix} = \begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} m_i \\ m_j \\ m_k \end{bmatrix}$$

The magnetic moment produced along the longitudinal axis is chosen so as to minimize the deviation of this axis from the earth's normal. This will

be discussed later. The magnetic moment components in the plane perpendicular to this direction are chosen such that:

$$(23) \qquad \begin{bmatrix} m_I \\ m_J \end{bmatrix} = C_1 \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} B_I \\ B_J \end{bmatrix}$$

where: $0 \leq \beta \leq \pi/2$.

C_1 is a parameter of proportionality and will be determined soon.

Because of its construction, the resistance of the coil perpendicular to the longitudinal axis of the satellite can be made much smaller than that of the coils aligned along this axis. When this is the case or when this coil has its own power supply, the torque required to rotate the satellite about its longitudinal axis is maximized by choosing $\beta = \pi/2$. The optimum solution to the control situation when these conditions do not exist is treated in Appendix II.

The relationship between the currents used to turn the satellite about its longitudinal axis and the earth's magnetic field for maximum torque is obtained from equations (21) and (22), with $\beta = \pi/2$, as:

$$(24) \qquad \begin{bmatrix} m_i \\ m_j \end{bmatrix} = C_1 \begin{bmatrix} -\sin\phi & \cos\phi \\ -\cos\phi & -\sin\phi \end{bmatrix} \begin{bmatrix} B_I \\ B_J \end{bmatrix}$$

with the assumption that the area, A_o, number of turns, N_o, and resistance R_o of the coils aligned along this axis are equal; the power required to produce the desired change in attitude about it is:

$$(25) \qquad P = \frac{R_o}{(N_o A_o)^2} (m_i^2 + m_j^2)$$

To minimize the time required for the change in attitude, the power is kept at its maximum value, P_m, during the maneuver. Using this value and equation (23) yields:

$$(26)\quad P_m = R_o \left(\frac{C_o}{N_o A_o} \right)^2 (B_I^2 + B_J^2)$$

in a form convenient for the analysis to follow with:

$$(27)\quad C_1 = \frac{C_o I_k}{L(t)}$$

where:

$$(28)\quad C_o = \pm \left(\frac{N_o A_o}{I_k} \right) \left(\frac{P_m}{R_o} \right)^{1/2} \quad \text{, a constant}$$

$$(29)\quad L(t) = \sqrt{B_I^2 + B_J^2}$$, where B_I and B_J depend on the orbit parameter time. The control equations which follow will be derived with C_o being positive when the satellite is to be rotated counterclockwise about its longitudinal axis and negative for clockwise rotation. Since the earth's magnetic fields vary along the orbit; L(t) will be a parameter of position and, hence, of time.

Using this interpretation of C_o and equations (20), (23), (27), (28) and (29), with $\beta = \pi/2$, yields:

$$(30)\quad \ddot{\phi}(t) = C_o L(t), \text{ when } \ddot{\phi}(t) \text{ and } \dot{\phi}(t) \text{ have the same sign}$$

(31) $\ddot{\phi}(t) = -C_o L(t)$, when $\ddot{\phi}(t)$ and $\dot{\phi}(t)$ have opposite signs.

At the beginning of the control period $t = t_1$, $\ddot{\phi}(t_1) = \dot{\phi}(t_1) = 0$.

The currents generating m_i and m_j are turned on with polarities so as to produce the required sense of rotation. At time $t = t_2$ the current polarities are reversed and applied until $t = t_3$, when $\dot{\phi}(t_3) = 0$.

Using this control procedure yields:

For $t_1 \leq t < t_2$:

(32) $$\ddot{\phi}(t) = C_o L(t)$$

(33) $$\dot{\phi}(t) = C_o \int_{t_1}^{t} L(\tau)\, d\tau$$

(34) $$\phi(t) = \phi(t_1) + C_o \int_{t_1}^{t} (t - \tau) L(\tau)\, d\tau$$

For $t_2 < t \leq t_3$:

(35) $$\ddot{\phi}(t) = -C_o L(t)$$

(36) $$\dot{\phi}(t) = C_o \int_{t}^{t_3} L(\tau)\, d\tau$$

(37) $$\phi(t) = \phi(t_1) + C_o \int_{t_1}^{t_2} (t_2 - \tau) L(\tau)\, dt + C_o \int_{t_2}^{t_3} (\tau - t_2) L(\tau)\, d\tau + C_o \int_{t}^{t_3} (t - \tau) L(\tau)\, d\tau$$

At $t = t_2$:

$$\int_{t_1}^{t_2} L(\tau)\, d\tau = \int_{t_2}^{t_3} L(\tau)\, d\tau$$

Letting:

(38) $\ell(t) = \int_{t_o}^{t} L(\tau)\, d\tau$, with t_o chosen for convenience in computation, one may determine t_2 given t_1 and t_3 from the relationship:

(39) $$\ell(t_2) = \ell(t_1) + \ell(t_3)$$

The period during which the satellite is visible to an observer at a ground station is determined by the location of the ground station and the satellite's orbit. The maximum period during which the satellite can be controlled by line-of-sight-transmission is less than the period of visibility, being limited by the increase in signal-to-noise ratio and acquisition difficulties near the horizon. t_1 and t_3, the times corresponding to the beginning and end of the control period, must be chosen within this maximum control period. The analysis which permits the minimum value of t_1 and the maximum value of t_3 to be computed is presented in Appendix I.

Since the magnetic field varies along the orbit, the response of the system will be time dependent. This is indicated by equation (29). The earth's magnetic field is assumed to be that of a dipole field whose north pole lies 70.1° west of Greenwich at 78.6° north latitude and whose south pole is 109.9° east of Greenwich at 78.6° south latitude. Except for low altitude orbits, this model is quite accurate.

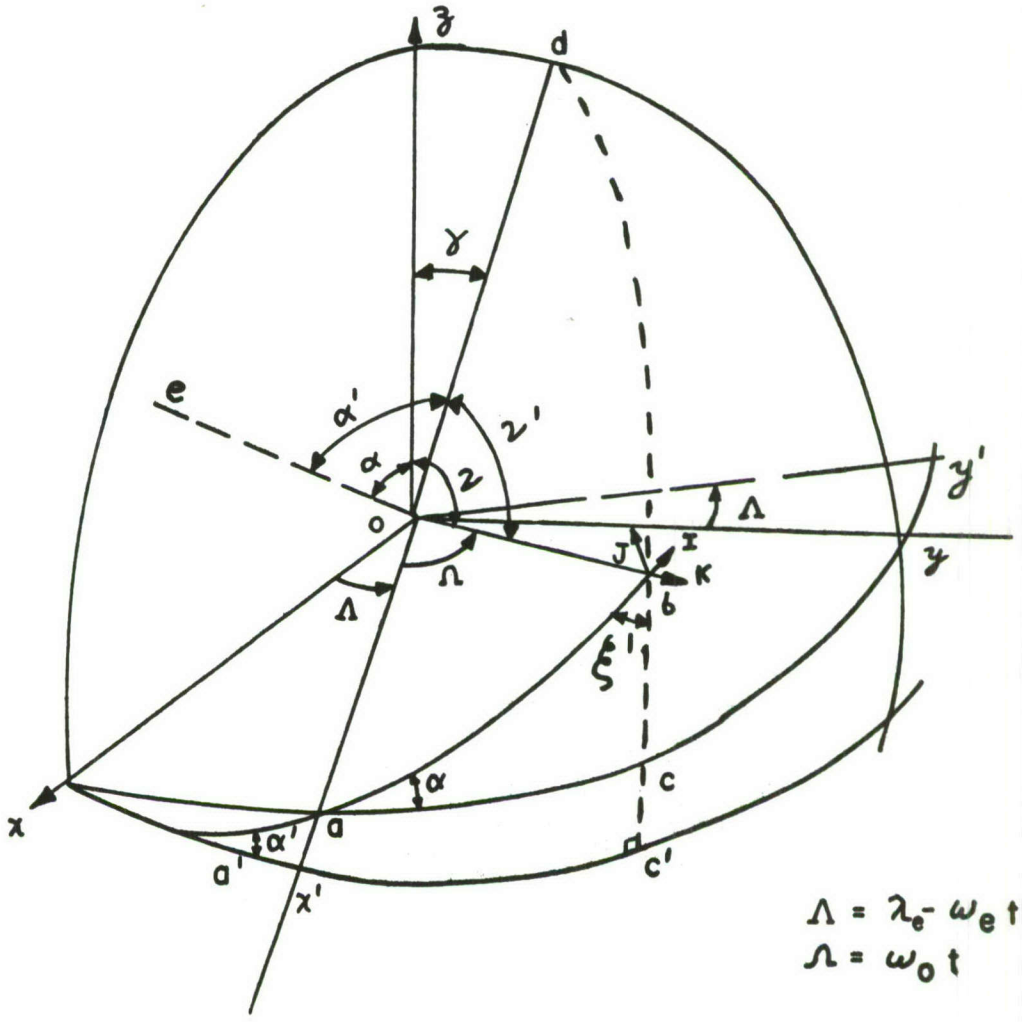


FIGURE 3. COORDINATES OF SATELLITE OF SATELLITE ORBIT

The orbit parameters which are to be used in the analysis that follows are indicated in Figure 3. Orbit time is taken to be zero at the equatorial crossing at which the angle α is defined. Referring to Figure 3, the earth's magnetic field at point b, the location of the satellite, is:

$$(40)\quad \begin{bmatrix} B_I \\ B_J \\ B_K \end{bmatrix} = B_o \begin{bmatrix} -\sin\psi' & \cos\xi' \\ -\sin\psi' & \sin\xi' \\ 2\cos\psi' & \end{bmatrix}$$

where: $B_o = 3.12 \times 10^{-5} \gamma_o^{-3}$ (weber/meters^2)

(As has been indicated, γ_o is the orbit radius in earth radii.)

From equations (29) and (40):

$$L_o(t) = B_o \sin\psi'$$

From geometric consideration this may be written as:

$$(41)\quad L(t) = B_o \sqrt{1 - (\sin\omega_o t \sin\alpha \cos\gamma + \tfrac{1}{2}\sigma_1(t)\sin\gamma)^2}$$

where

$$(42)\quad \sigma_1(t) = [1+\cos\alpha]\sin(\omega_o t - \omega_e t + \lambda_e) + [1-\cos\alpha]\sin(\omega_o t + \omega_o t - \lambda_e)$$

Besides time, t, the variable parameters in this expression are the angle between the equatorial and orbital plane, α ; the equatorial crossing angle, λ_e ; and the orbital radius, γ_o . From the analysis which has been

presented, the orientation of the satellite in a circular orbit is obtained using these parameters and the times the coil currents are turned on, reversed in polarity and turned off.

The above analysis has been applied to a satellite in a circular orbit which is 700 nautical miles above the earth and inclined 40° to the equatorial plane for various equatorial angle crossings. This satellite is of particular interest to the Communications Branch at the Rome Air Development Center. The results of the analysis are presented in Section 6.

5. THE CONTROL REQUIRED TO MINIMIZE CROSS-COUPLED TORQUES

When the satellite is rotated about its longitudinal axis, a cross-coupled torque is produced normal to this direction. This was indicated in equation (9). The angular velocity and position are controlled by varying m_i and m_j. m_k is used to minimize the effect of the cross-coupled torque. It is possible to select m_k so as to cancel either the component of torque along the orbital direction or along the normal to orbital plane, but not both. An optimum control is to vary m_k so that the time response of the two components of cross-coupled torque at the end of the control period will be in phase quadrature and of equal magnitude. This would minimize their disturbance since this net torque is the root-mean-square sum of its two components. It will be shown subsequently that by selecting

$$(43) \qquad m_k \doteq m_K = \left(\frac{B_K}{B_J}\right) m_J$$

The components of cross-coupled torque are to a first order approximation in phase quadrature and of the same order of magnitude.

Substituting equation (43) into equations (18) and (19) yields:

$$\ddot{\psi} + a_0 H_0 \dot{\psi} + a_0 \psi = -\left(\frac{I_k}{I_0}\right) \omega_0 \dot{\phi} \tag{44}$$

$$\ddot{\theta} + a_0 H_0 \dot{\theta} + a_0 \theta = -\left(\frac{I_k}{I_0}\right)\left(\frac{B_K}{B_J}\right) \ddot{\Phi} \tag{45}$$

Using equations (40) and (42) and the geometrical relationship among the parameters of figure 3, it is found that:

$$\frac{B_K}{B_J} = \frac{2 \sin\alpha \sin\omega_0 t + \nabla_i(t) \tan\gamma}{\cos(\lambda_e - \omega_e t) \tan\gamma \sin\alpha - \cos\alpha} \tag{46}$$

Substituting equations (32), (33), (35), (36) and (46) into equations (44) and (45) yields:

For $t_1 \leq t < t_2$:

$$\ddot{\psi} + a_0 H_0 \dot{\psi} + a_0 \psi = -\left(\frac{I_k}{I_0}\right) \omega_0 C_0 \int_{t_1}^{t} L(\tau)\, d\tau \tag{47}$$

$$\ddot{\theta} + a_0 H_0 \dot{\theta} + a_0 \theta = \left(\frac{I_k}{I_0}\right)\left(\frac{2 \sin\alpha \sin\omega_0 t + \nabla_i(t) \tan\gamma}{\cos(\lambda_e - \omega_e t) \tan\gamma \sin\alpha - \cos\alpha}\right) C_0 L(t) \tag{48}$$

For $t_2 < t \leq t_3$:

$$\ddot{\psi} + a_0 H_0 \dot{\psi} + a_0 \psi = \left(\frac{I_k}{I_0}\right) \omega_0 C_0 \int_{t}^{t_3} L(\tau)\, d\tau \tag{49}$$

$$(50)\quad \ddot{\theta} + a_o H_o \dot{\theta} + a_o \theta = \left(\frac{I_A}{I_o}\right)\left(\frac{2\sin\alpha\sin\omega_o t + \sigma_1(t)\tan\gamma}{\cos(\lambda_e - \omega_e t)\tan\gamma\sin\alpha - \cos\alpha}\right) C_o L(t)$$

Solutions to these equations obtained for the satellite problem of interest to RADC are indicated in Appendix II.

To show that selecting m_k according to equation (43) will produce the desired minimization of cross-coupled torque, it will be necessary to simply if the results indicated in equations (47), (48), (49) and (50) by using certain approximations.

Since $\dot{\phi}(t)$ is continuous for $t_1 \leq t \leq t_3$, with $\dot{\phi}(t_1) = \dot{\phi}(t_3) = 0$ it may be represented by a Fourier sine series over this interval. Furthermore, this series can be differentiated term by term to represent $\ddot{\phi}(t)$ over the interval except at its point of discontinuity, $t = t_2$. This result is:

$$(51)\quad \dot{\phi}(t) = \sum_{n=1}^{\infty} b_n \sin[n\omega_1(t-t_1)]$$

$$(52)\quad \ddot{\phi}(t) = \omega_1 \sum_{n=1}^{\infty} n\, b_n \cos[n\omega_1(t-t_1)]$$

where:

$$b_n = \frac{1}{2\pi n}\int_{t_1}^{t_3} \ddot{\phi}(t)\cos[n\omega_1(t-t_1)]\,dt$$

(53) $$\omega_1 = \frac{\pi}{(t_3 - t_1)}$$

on using equations (32 and (35):

(54) $$b_m = \frac{C_0}{2m\pi}\left\{\int_{t_1}^{t_2} L(t)\cos n\omega_1(t-t_1)\,dt - \int_{t_2}^{t_3} L(t)\cos n\omega_1(t-t_1)\,dt\right\}$$

These coefficients are approximately proportional to n^{-2} . This can be seen by assuming that L(t) = L(t_2), a constant, over the control period. For this condition $t_2 = \frac{1}{2}(t_1 + t_3)$ and:

$$b_n = \begin{cases} \dfrac{\omega_1 C_0 L(t_2)}{n^2\pi} & , \ n \text{ odd} \\ 0 & , \ n \text{ even} \end{cases}$$

For orbits for $|\tan\alpha| \leq 2$, orbits which do not pass close to the magnetic poles:

$$\frac{B_K}{B_J} \doteq -2\tan\alpha \sin(\omega_0 t)$$

For high and medium altitude orbits $0 \leq \frac{\omega_0}{\omega_e} \leq \frac{1}{3}$. Hence, we may assume that (B_K/B_J) varies slowly compared to $\dot{\phi}(t)$ and $\ddot{\phi}(t)$. Setting (θ_K/θ_J) equal to its value at the center of the control period one obtains the approximation:

(55) $$\left(\frac{B_K}{B_J}\right) \doteq -2\tan\alpha \sin(\omega_0 t_2)$$

The amount of gravity gradient damping that can be obtained at the present state of technology is small. Because of this, its effect can be

neglected during the control period and the period immediately following the time the coil currents are turned off, $t = t_3$.

Using equations (51) to (55) and the modification suggested above, one obtains from the relationships indicated by equations (44) and (45) that for $t_1 \leq t \leq t_3$:

$$(56) \quad \ddot{\psi} + a_0 \psi = -\omega_0 \left(\frac{I_k}{I_0}\right) \sum_{n=1}^{\infty} b_n \sin\left[n\omega_1 (t-t_1)\right]$$

$$(57) \quad \ddot{\theta} + a_0 \theta = 2\omega_1 \sin(\omega_0 t_2) \tan\alpha \left(\frac{I_k}{I_0}\right) \sum_{n=1}^{\infty} n b_n \cos\left[n\omega_1 (t-t_1)\right]$$

Solving these differential equations yields:

$$(58) \quad \psi(t) = -\omega_0 \left(\frac{I_k}{I_0}\right) \sum_{n=1}^{\infty} \frac{b_n}{n^2\omega_1^2 - a_0} \left\{ \frac{n\omega_1}{\sqrt{a_0}} \sin\left[\sqrt{a_0}(t-t_1)\right] - \sin\left[\omega_1 (t-t_1)\right] \right\}$$

$$(59) \quad \theta(t) = 2\omega_1 \sin(\omega_0 t_2) \tan\alpha \left(\frac{I_k}{I_0}\right) \sum_{n=1}^{\infty} \frac{n b_n}{(n^2\omega_1^2 - a_0)} \left\{ \cos\left[\sqrt{a_0}(t-t_1)\right] - \cos\left[n\omega_1 (t-t_1)\right] \right\}$$

At the end of the control period, $t = t_3$:

$$(60)\quad \psi(t_3) = -\left(\frac{I_k \omega_0 \omega_1}{I_0 \sqrt{a_0}}\right) \sum_{n=1}^{\infty} \frac{n b_n \sin(\pi \sqrt{a_0}/\omega_1)}{(n^2\omega_1^2 - a_0)}$$

$$(61)\quad \dot{\psi}(t_3) = -\left(\frac{I_k \omega_0 \omega_1}{I_0}\right) \sum_{n=1}^{\infty} \frac{n b_n}{(n^2\omega_1^2 - a_0)} \left[\cos(\pi \sqrt{a_0}/\omega_1) - (-1)^n\right]$$

$$(62)\quad \theta(t_3) = 2 \sin(\omega_0 t_2) \tan\alpha \left(\frac{I_k \omega_1}{I_0}\right) \sum_{n=1}^{\infty} \frac{n b_n}{(n^2\omega_1^2 - a_0)} \left[\cos \frac{\pi \sqrt{a_0}}{\omega_1} - (-1)^n\right]$$

$$(63)\quad \dot{\theta}(t_3) = -2 \sin(\omega_0 t_2) \tan\alpha \left(\frac{I_k \omega_1 \omega_2}{I_0}\right) \sum_{n=1}^{\infty}$$

$$\frac{n b_n}{n^2\omega_1^2 - a_0} \sin\left(\frac{\pi \sqrt{a_0}}{\omega_1}\right)$$

After the control currents are turned off, the satellite will oscillate about the I and the J axes. This response is:

$$(64)\quad \psi(t) = \sqrt{\left[\psi(t_3)\right]^2 + \left[\dot{\psi}(t_3)/\sqrt{a_0}\right]^2}\,\sin\left[\sqrt{a_0}(t-t_3) - \tan^{-1}\left(\frac{\dot{\psi}(t_3)}{\sqrt{a_0}\,\psi(t_3)}\right)\right],$$

$$(65)\quad \theta(t) = \sqrt{\left[\theta(t_3)\right]^2 + \left[\dot{\theta}(t_3)/\sqrt{a_0}\right]^2}\,\cos\left[\sqrt{a_0}(t-t_3) - \tan\left(\frac{-\sqrt{a_0}\,\theta(t_3)}{\dot{\theta}(t_3)}\right)\right],$$

for $t \geq t_3$:

From equations (60) to (63) it is seen that:

$$(66)\quad \tan^{-1}\frac{\dot{\psi}(t_3)}{\sqrt{a_0}\,\psi(t_3)} = \tan^{-1}\left(\frac{-\sqrt{a_0}\,\theta(t_3)}{\dot{\theta}(t_3)}\right),$$

This, together with the formulation of equations (65) and (66), shows that $\psi(t)$ and $\theta(t)$ are, and remain, in phase quadrate for $t \geq t_3$.

By comparing the amplitudes of $\psi(t)$ and $\theta(t)$, it is also seen that they are of the same order of magnitude.

In the dynamics indicated by equations (65), (66) and (67), the longitudinal axis of the satellite moves on the surface of an elliptic cone, whose axis is the local normal to the earth, with uniform angular velocity about the axis of $\sqrt{a_0}$. This may be represented as the combination of nutation and precession motions with equal periodicity.

Including the effect of gravity gradient damping in the dynamics would show that the major and minor axes of this conic would decay exponentially with a time constant of $\frac{1}{2} a_0 H_0$. It would also cause the phase relationship between ψ and θ to differ slightly from being one of quadrature and reduce the frequency of oscillations by a small amount. Its effect would increase the ratio of the nutation to the precession amplitudes by a small amount. The variation of the ratio (B_K / B_J) over the control period produces a nutation about the orbital direction (J -) axis which persists after the coil currents are turned off. This effect is superimposed on the other motion and occurs when ψ and θ are small, with the same oscillation frequency and decay rate as the natural mode. For control periods in which (B_K / B_J) is reasonably constant, this effect can be neglected.

The means of extending this control technique to the situation, where the longitudinal axis of the satellite deviates by a significantly large angle from the local normal to the earth, is provided by the concepts indicated above. The condition that the satellite be turned about its longitudinal axis in the minimum amount of time would continue to be imposed.

In addition, it would be required that the ratio of the energy going into nutation to that going into precession be minimized. After the control currents are turned off, the satellite will oscillate in a manner similar to that of a heavy top. The frequencies of precession and nutation would now differ.

6. RADC APPLICATION

The results of this report have been applied to a satellite in a circular orbit 700 nautical miles above the earth, which is of particular interest to RADC. The constants of this computation are:

As defined in Figure 3:

$$\alpha = 40.00^\circ$$

$$\gamma = 11.40^\circ$$

$$\omega_o = 9.434 \times 10^{-4}$$

$$\omega_e = \frac{2\pi}{(24)(60)(60)} = 7.272 \times 10^{-5}$$

As defined in Figures I-1 and I-2 of Appendix I, using the coordinates of RADC's Floyd site:

$$\lambda_s = 7.000^\circ$$

$$\gamma_f = 46.80^\circ$$

$$\lambda_f = 84.76^\circ$$

As defined in Equations (10), (12) and (21):

$$H_o = 12.72$$

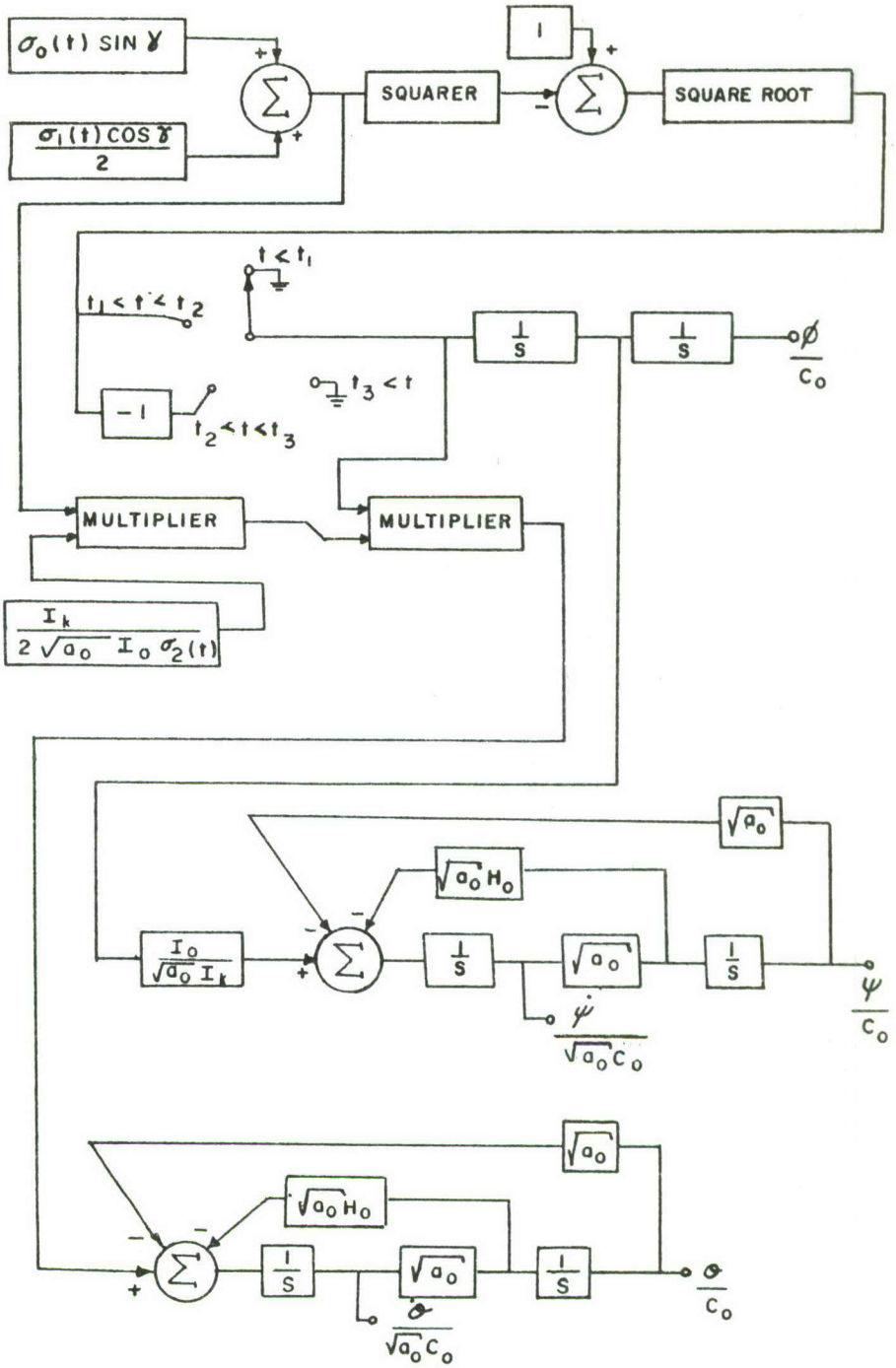
$$a_o = \frac{K_o}{\gamma_o^3}\left(1 - \frac{I_{\phi\phi}}{I_o}\right) = 2.474 \times 10^{-6}$$

As is indicated by equation (28), C_o depends on the satellite's moment of inertia about its longitudinal axis, the maximum power supply output and the details of coil construction. In order to permit these factors to be used as design parameters, the final results have been made independent of C_o. Because of the way that C_o enters into the computation, this may be done by dividing by it.

Using Equation (I-5), minimum values of t_1 and maximum values of t_3 were determined over the range of λ_e applicable to the Floyd site. Using Equation (39), values of t_2 are determined as a function of λ_e and t_3. This is done by setting t_1 at its minimum value and varying t_3 between this value and its maximum. A Univac 1103A digital computer was used to determine the values of t_1, t_2 and t_3.

The change in the satellite's attitude about its spin axis was determined from Equation (37), using the computed values of t_1, t_2 and t_3. This was done using Reeves 400 and 500 analogue computers in parallel. The block diagram of the analogue computer circuit is indicated in Figure 4. The result of each set of computations is a characteristic of constant λ_e. The final result, a graph of characteristics that permit the change in attitude to be determined as a function of the control period and λ_e, is shown in Figure 5. Using these characteristics permits one to determine the times that coil currents must be turned on, reversed in polarity, and turned off in order to optimally produce a specified change in attitude for a satellite in an orbit having a given λ_e.

It was shown in the report that after the coil currents are turned off, the satellite will deviate from alignment with the local normal to the earth. It was shown in Equations (64) and (65) that $\theta(t)$ and $\psi(t)$ are bounded



$\frac{1}{s}$ REPRESENTS INTEGRATION

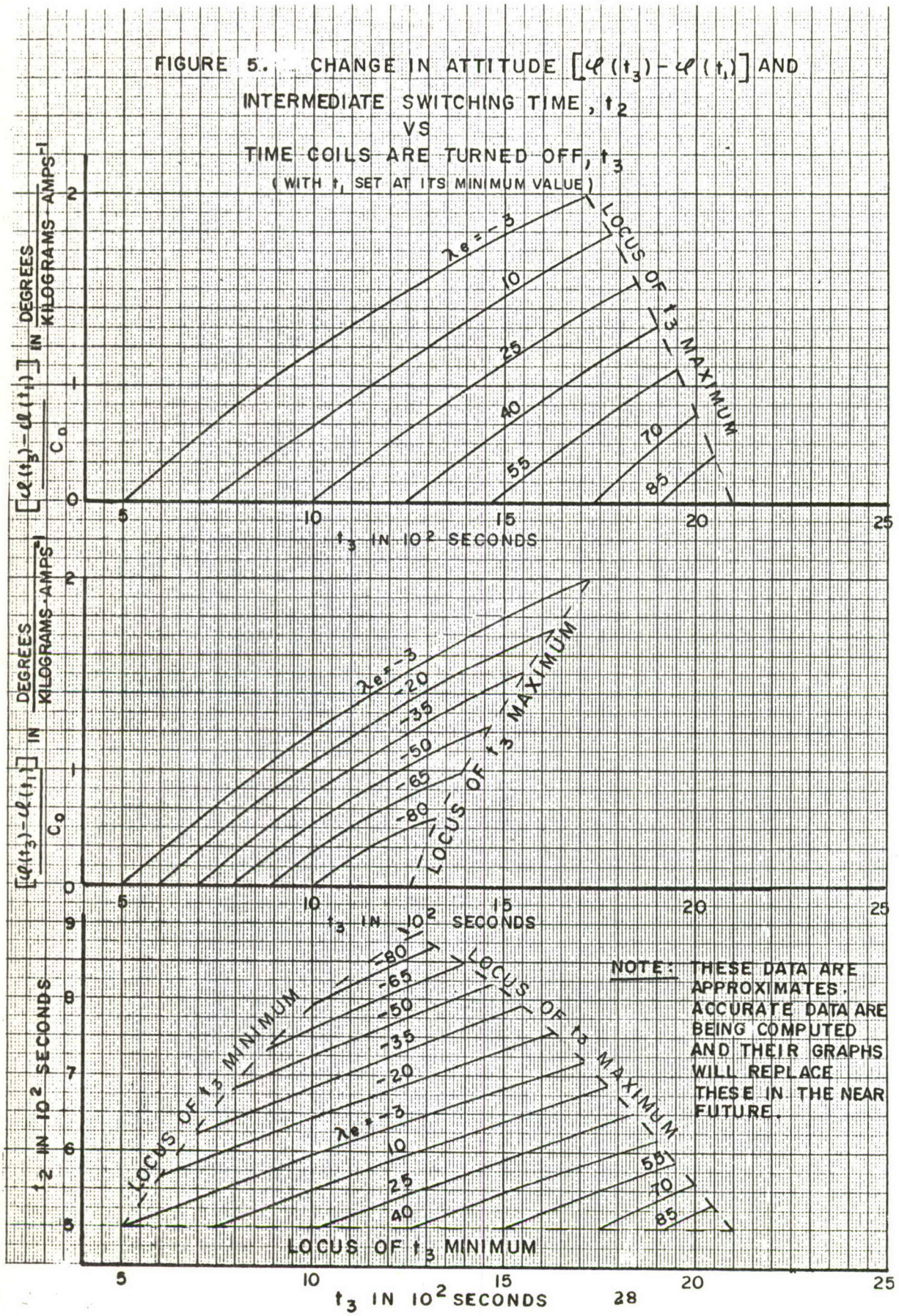
$\sigma_0(t) = \text{SIN}\ \alpha\ \text{SIN}\ \omega_0 t$

$\sigma_1(t) = (1 + \cos\alpha)\ \sin\left[(\omega_0 - \omega_e)t + \lambda_e\right] + (1 - \cos\alpha)\ \sin\left[(\omega_0 + \omega_e)t - \lambda_e\right]$

$\sigma_2(t) = \cos\alpha\cos\gamma - \sin\alpha\sin\gamma\cos(\lambda_e - \omega_e t)$

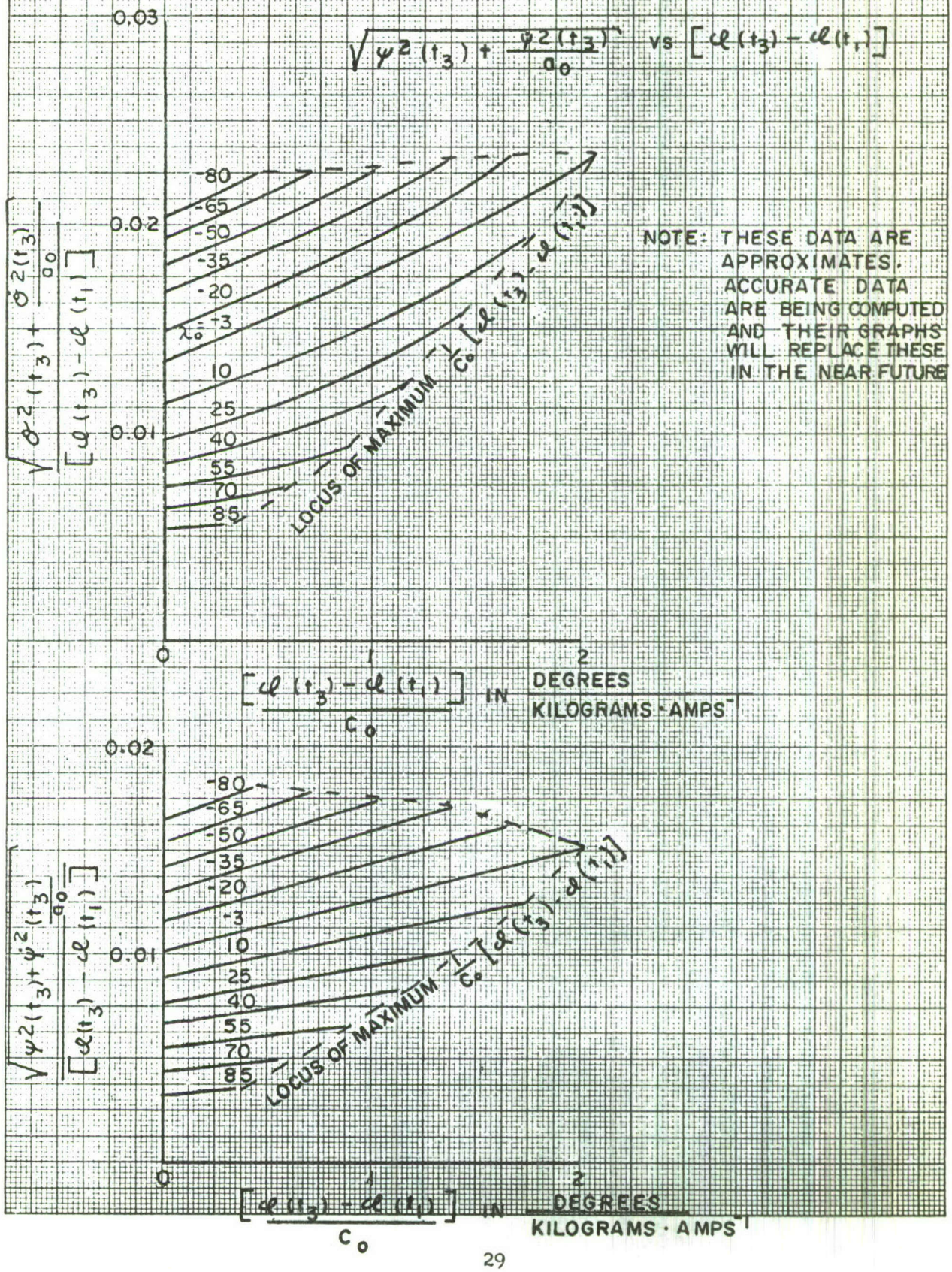
FIGURE 4. ANALOGUE COMPUTER CIRCUIT

FIGURE 5. CHANGE IN ATTITUDE $[\varphi(t_3) - \varphi(t_1)]$ AND INTERMEDIATE SWITCHING TIME, t_2 VS TIME COILS ARE TURNED OFF, t_3 (WITH t_1 SET AT ITS MINIMUM VALUE)



NOTE: THESE DATA ARE APPROXIMATES. ACCURATE DATA ARE BEING COMPUTED AND THEIR GRAPHS WILL REPLACE THESE IN THE NEAR FUTURE.

FIGURE 6. COMPUTATION OF $\sqrt{\phi^2(t_3) + \frac{\dot{\phi}^2(t_3)}{a_0}}$ AND $\sqrt{\psi^2(t_3) + \frac{\dot{\psi}^2(t_3)}{a_0}}$ VS $\left[\ell(t_3) - \ell(t_1)\right]$



by $\sqrt{\theta^2(t_3) + \frac{\dot{\theta}^2(t_3)}{a_0}}$ and $\sqrt{\psi^2(t_3) + \frac{\dot{\psi}^2(t_3)}{a_0}}$ respectively in this period (neglecting the perturbative effect due to variations in the ratio (B_K/B_T) during the control period). These quantities were determined for the various value of $[\phi(t_3) - \phi(t_1)]$ and were computed simultaneously with them. The required analogue computer circuitry is shown in Figure 4. The results are presented in Figure 6.

7. CONCLUDING REMARKS

The technique which has been developed may be applied to a broad class of control problems. It is applicable to positioning a solar sail, an antenna array, or any device which is an integral part of the satellite. The method requires that the satelites moment of inertia about its longitudinal axis be an order of magnitude less than those about axes normal to it, and that these other moments of inertia be approximately equal. It is effective for any orbit except those passing close to the magnetic poles.

The technique permits the satellites attitude about its longitudinal axes to be changed in a minimum amount of time with the additional condition that the angular velocity about this axes return to its initial state of zero at the end of the control period. The other purpose of the control period is to minimize the deviation of the longitudinal axes from the local normal to the earth. This is accomplished by impossing a phase quadrature condition on the components of torque normal to this axis.

The analysis was applied to a satellite and a type of orbit which are of particular interest to the Rome Air Development Center. Changes of attitude about the longitudinal axes and deviation of this axes from the earth's normal were

determined as parameter of the coil constants and power supply output. These results were obtained over the range of control period and equatorial crossing angle, λ_e , which were applicable to the orbits as viewed from RADC Floyd site.

APPENDIX I

DETERMINING THE PERIOD

OVER WHICH THE SATELLITE CAN BE CONTROLLED

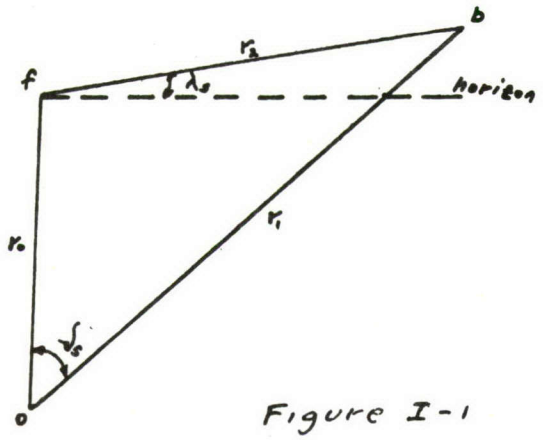


Figure I-1

Referring to Figure I-1 , let o be the center of the earth, f be a point on its surface, and b be the location of a satellite at an angle of λ_s above the horizon of point f. From geometrical considerations, one may write:

(I-1) $$r_1^2 = r_0^2 + r_2^2 + 2 r_0 r_2 \sin \lambda_s$$

(I-2) $$r_2^2 = r_1^2 + r_0^2 - 2 r_0 r_1 \cos \gamma_s$$

Using these equations and $\gamma_0 = (r_1/r_0)$, the condition that the satellite will be at least λ_s degrees above the horizon at point f is that:

(I-3) $$\gamma_0 \cos \gamma_s \geq \cos^2 \lambda_s + \lambda_s \sqrt{\gamma_0^2 - \cos^2 \lambda_s}$$

The satellite orbit is shown in Figure I-2 as arc ab. The earth's equator lies in x-y plane and its north pole in the y-z plane. (This is done to make the system consistent with that used elsewhere in this report.)

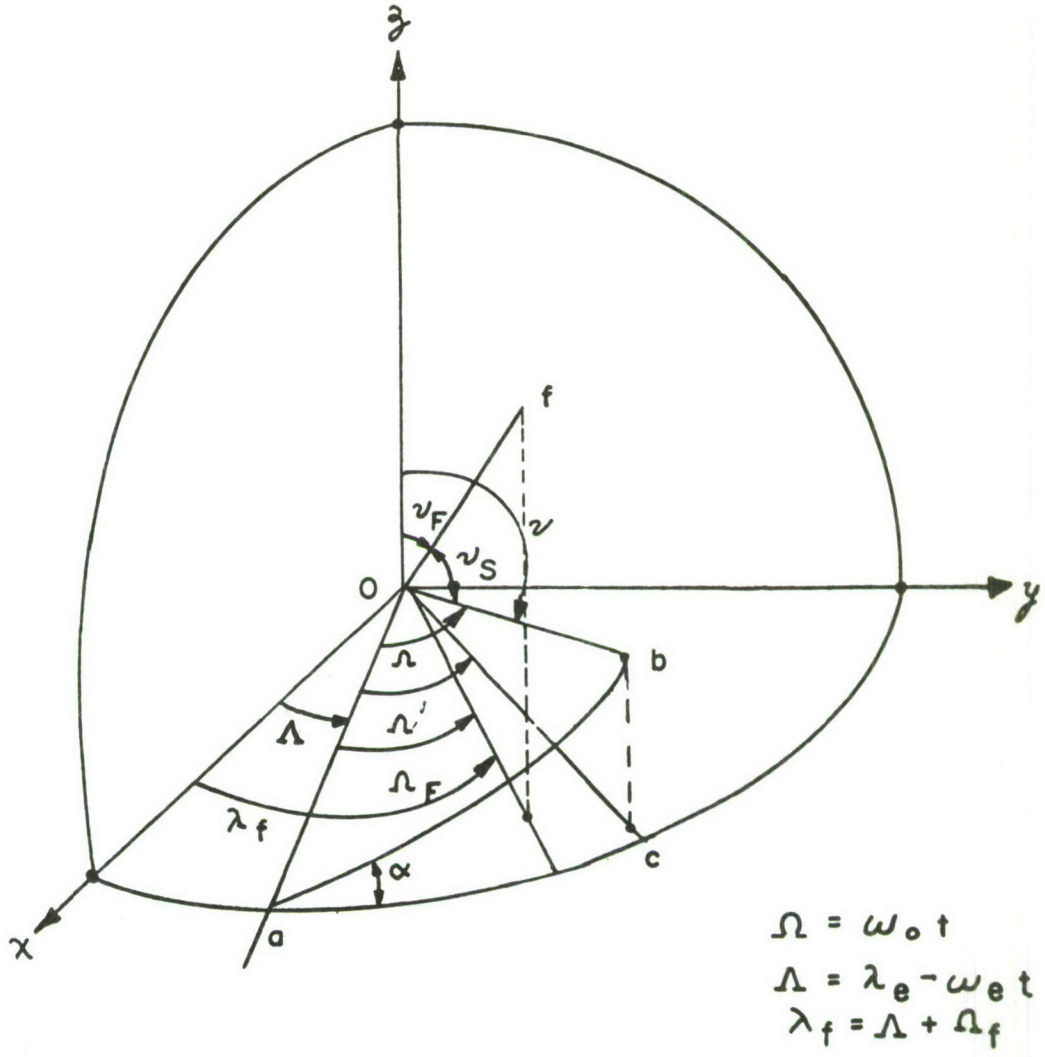


FIGURE (I-2) COORDINATES OF SATELLITE ORBIT & GROUND STATION

At orbit time $t = 0$, the satellite crosses the equator making an angle λ_e with the x axis. The following analysis is made with the satellite at point b, from geometrical considerations:

$$\cos \nu_s = (\cos \Omega' \sin \nu) \cos \Omega_f \sin \nu_f + (\sin \Omega' \sin \nu) \sin \Omega_f \sin \nu_f + (\cos \nu) \cos \nu_f$$

$$= (\cos \Omega) \cos \Omega_f \sin \nu_f + (\sin \Omega \cos \alpha) \sin \Omega_f \sin \nu_f + (\sin \Omega \sin \alpha) \cos \nu_f$$

with:

$$\Omega = \omega_e t$$

$$\Omega_f = \lambda_f - \Lambda$$

$$\Lambda = \lambda_e - \omega_e t$$

This can be written as:

$$\text{(I-4)} \quad \cos \nu_s = \cos \omega_o t \left[\sin \nu_f \cos (\lambda_f - \lambda_e + \omega_e t)\right] + \sin \omega_o t \left[\cos \alpha \sin \nu_f \sin (\lambda_f - \lambda_e + \omega_e t) + \sin \alpha \cos \nu_f\right]$$

Assuming that the satellite's attitude can be controlled when it is

within a cone of central angle $\frac{\pi}{2} - \lambda_s$ whose apex is at point f, and using equation (I-3) and (I-4), the satellite can be controlled by line of sight transmission for all time, t, such that:

$$(I\text{-}5) \quad \gamma_o \cos \omega_o t \left[\sin \gamma_f \cos (\lambda_f - \lambda_o + \omega_e t) \right]$$

$$+ \gamma_o \sin \omega_o t \left[\cos \alpha \sin \gamma_f \sin (\lambda_f - \lambda_o t + \omega_e t) + \sin \alpha \cos \gamma_f \right]$$

$$\geq \cos^2 \lambda_s + \sin \lambda_s \sqrt{\gamma_o^2 - \cos^2 \lambda_s}$$

APPENDIX II

REVISION IN CONTROL PROCEDURE REQUIRED WHEN ALL COIL CURRENTS ARE PROVIDED BY ONE POWER SUPPLY

The optimization procedure is modified when the power supply used to produce m_i and m_j is also used to produce m_k. The angle between the components of magnetic moment and earth magnetic field vectors in a plane transverse to the satellite's longitudinal axis is varied with orbit position, rather than be maintained constant at This modification will be derived.

Neglecting back e. m. f. and inductive effects the power required to produce the magnetic moments is:

$$P = R_i\left(\frac{m_i}{N_i A_i}\right)^2 + R_j\left(\frac{m_j}{N_j A_j}\right)^2 + R_k\left(\frac{m_k}{N_k A_k}\right) \tag{II-1}$$

where: the N's, A's and R's represent the number of turns, area and resistance of the coils.

Let the coils be constructed such that:

$$\frac{N_o^2 A_o^2}{R_o} = \frac{N_i^2 A_i^2}{R_i} = \frac{N_j^2 A_j^2}{R_j} = k_o\left(\frac{N_k^2 A_k^2}{R_k}\right) \tag{II-2}$$

and the power supply be set at its maximum value, Pm, then using equation (28), one may write:

$$C_o^2 I_k^2 = m_i^2 + m_j^2 + k_o m_k^2 \tag{II-3}$$

From Equations (20), (22), (23) and (49) one obtains:

$$I_k \dot{\phi} = C_1 L^2(t) \sin\beta$$

$$I_k C_0 = C_1 \left\{ L^2(t) + k_0 B_K^2 \left[\cos\beta - \left(\frac{B_I}{B_J} \right) \sin\beta \right]^2 \right\}^{1/2}$$

From which

$$\dot{\phi}^{-2} = C_0^2 L^2(t) \sin^2\beta \left[1 + k_0 \delta_2 \sin^2\beta (\cot\beta - \delta_1)^2 \right]^{-1} \tag{II-4}$$

where:

$$\delta_1 = \frac{B_I}{B_J} \tag{II-5}$$

$$\delta_2 = \frac{B_K^2}{L^2(t)} \tag{II-6}$$

Differentiating Equation (II-4) with respect to β and setting the result equal to zero yields:

$$\sin^2\beta \left[\cos\beta (1 + k_0 \delta_2) - k_0 \delta_2 \delta_1 \right] = 0$$

This equality will be satisfied for:

$$\beta = \pm n\pi \qquad \text{and}$$

$$\beta = \tan^{-1} \left[\delta_1 (1 + k_0^{-1} \delta_2^{-1}) \right] \pm n\pi$$

$n = 0, 1, 2, \ldots$

By physical reasoning, $\dot{\phi}^2$ will be a minimum (zero) for the first of these conditions and a maximum for the second. Selecting $\beta < \pi$, to be consistent with the definition of C_o, results in choosing:

$$\text{(II-7)} \qquad \beta = \tan^{-1}\left[\delta_1^{-1}\left(1 + k_o^{-1}\delta_2^{-1}\right)\right]$$

Using the analysis presented by Equations (20) to (29) shows that the required modification in control procedure is obtained by replacing

$L(t)$

by

$L(t) \sin\beta$

The necessary modification factor $\sin\beta$ can be expressed as a parameter of orbit t, using equations (40), (II-5), (II-6) and (II-7). This result is:

$$\text{(II-8)} \qquad \sin\beta = \left[1 + \frac{4k_o\sigma_2^2\left[1 - \frac{L(t)}{B_o}\right]}{\left[\sigma_2(t) - \frac{L^2(t)}{B_o^2}\right]\left\{1 + (4k_o - 1)\left[1 - \frac{L(t)}{B_o}\right]^2\right\}}\right]^{-\frac{1}{2}}$$

where: $\sigma_2(t) = \cos\alpha\cos\gamma - \sin\alpha\sin\gamma\cos(\lambda_e - \omega_e t)$

$L(t)$ is given by equation (41).

GLOSSARY OF SYMBOLS

Symbol	Description	Defined on Page
A	Angular velocity transformation matrix	3
A_o, N_o, R_o	Area, number of turns and resistance of a coil aligned along the satellite's longitudinal axis	11
a_o	Constant of proportionality in the equations of motion	10
B_o	Proportionality factor of the earth's magnetic field and the satellite altitude	16
B_i, B_j, B_k, B_I^i, B_J^j, B_K^k	Components of the earth's magnetic field in body-fixed and orbit-fixed coordinates	9
b_n	Fourier sine series coefficient	19
C_o	Constant of proportionality of the satellite's moment of inertia, coil construction, and power supply output	12
C_1	Parameter of proportionality between the components of the magnetic moments and those of the earth's magnetic field	11
E	Euler transformation matrix	3
H_o	Viscous damping constant	
I_o	Satellite's moment of inertia about an axis in i-j plane	7
I_i, I_j, I_k I_I^i, I_J^j, I_K^k	Satellite's moment of inertia in body-fixed and orbit-fixed coordinates	6, 10
I, J, K	Orbit-fixed coordinate axes	3
i, j, k	Body-fixed coordinate axes	3
K_o	Constant of proportionality of the gravity gradient torque	7
k_o	Proportionality constant relating the area, number of turns, and resistance of a coil aligned along the longitudinal axis of the satellite to those of a coil normal to it.	36
$L(t)$	Magnitude of the earth's magnetic field normal to the earth's radius (K-axis) at the satellite's orbit position	12

Symbol	Description	Defined on Page
$I(t)$	Integral of $L(t)$	14
m_i, m_j, m_k, m_I, m_J, m_K	Magnetic moment components in body-fixed and orbit-fixed coordinates	9
P, P_m	Instantaneous and maximum values of the power dissipated in the coils	11, 12
r_o, r_1, r_2	Earth's radius and satellite's distances from the center of the earth and the transmitter site	7, 32
T_i, T_j, T_k	Mechanical torques about the body-fixed axes	6
T_{Bi}, T_{Bj}, T_{Bk}	Torques about the body-fixed axes due to the interaction of the satellite's magnetic moment and the earth's magnetic field	8
T_{gi}, T_{gj}, T_{gk}, T_{ui}, T_{uj}, T_{uk}	Damped and Undamped gravity gradient torque about the body-fixed axes	7, 8
t_1, t_2, t_3	Times the coils are switched on, reversed in polarity; and turned off ($t = 0$ when satellite crosses equatorial plane)	13
β	Angle between the magnetic moment and earth's field vectors in the I-J plane	11
δ_1, δ_2	Earth's magnetic field ratios	37
λ_e	Angle between a reference (160.1° west of Greenwich longitude) and satellites orbit at equatorial plane	15
γ_o	Orbit radius in earth radii	7
$\sigma_1(t)$, $\sigma_2(t)$	Orbital-magnetic field parameters	17, 38
ω_o	Angular velocity of the orbital reference frame with respect to a frame fixed at the earth's center	5
ω_1	Control period frequency	20
ω_i, ω_j, ω_k	Angular velocities about the body-fixed axes	3
φ, θ, ψ	Euler angles	4

Symbol	Description	Defined on Page
λ_s, ν_s, λ_f, Ω_f, ν_f, Ω'	Orbital and ground station parameters	26, 27
α, α', ν, ν', Ω, Λ	Orbital parameters	15

ORAL PRESENTATION OUTLINE

I Background of the Problem

A. Advantages of a Pseudo Passive Satellite Network (Slide #1)

B. Method of Positioning a Satellite by Controlling the Attitude of a Solar Sail (Slide #2)

II The Proposed Attitude Control System

A. Outline of the Technique

1. Conditions under which technique can be applied

2. Its advantages over convention control systems

B. The Coordinate Systems used in the Problem and their Interrelationship (Slide #3)

C. Equations of Motion, Small Angle Approximations

1. Torque due to the satellite's motion (Slide #4)

2. Torque due to the gravity-gradient effect (Slide #5)

3. Torque due to the interaction of the magnetic moments produced by the satellites coils and the magnetic field of the Earth (Slide #6)

4. Resultant Equations of Motion (Slide #7)

D. Time Optimal Control Equation

1. The maximum Torque Criterion

a. Maximum power condition (Slide #8)

b. Maximum field interaction condition (Slide #9)

2. Method of controlling coil current

3. Model of the Earth's magnetic field (Slide #10)

4. Resultant control equation (Slide #11)

E. The Control Required to Minimize Cross-Coupled Torques

1. Additional condition empassed on coil currents (Slide #12)

2. Resultant differential equations (Slide #13)

3. Approximate solution of equations (Slide #14)

F. Analogue Computation

1. Simplified functional block diagram of computer circuit (Slide #15)

2. Computed results

a. Coil switching times (Slide #16)

b. Change in attitude about longitudinal axis of satellite (Slide #17)

c. Cross coupled disturbances (Slide #18)

G. Summary

1. Discussion of computed results

2. Means of extending analysis to situations where there are large disturbances torques present.

JAMES HARRY SCHARFF

MAJOR USAF

AIR FORCE WEAPONS LABORATORY

KIRTLAND AFB, NEW MEXICO

"LIGHTNING - AN AEROSPACE NEMESIS?"

UNCLASSIFIED

LIGHTNING

An Aerospace Nemesis

Introduction - It is extremely interesting to scientists that a natural phenomenon which has been with us since the beginnings of the universe still has unexplored characteristics. Even in this day of megatons and moon rockets, the phenomenon that Ben Franklin made famous is still a fruitful field for scientific endeavor. It was thus that the whims of nature, some three years ago, launched the Air Force into an extensive study of particular aspects of lightning.

Background - In the fall of 1961, a series of thunderstorms in Italy repeatedly caused severe damage to the Jupiter weapon system. The nature of the damage varied widely. There was the common effect on power systems and communications. In addition to this, however, there were many effects that were not so common. Elements of the weapon system that were supposedly isolated and immune to all external environments were seriously affected. For a while there was some concern that certain of these effects might involve the nuclear safety of weapons.

In December 1961, Ballistic Systems Division was assigned the responsibility for evaluating the effects of lightning on USAF missile systems. This responsibility was then further delegated to the Development Directorate of AFSWC, and now to the Analysis Division of the Air Force Weapons Laboratory. The work to date has been focused on two main goals. First, to obtaining a better understanding of the natural phenomenon of lightning, and second, to developing techniques for the simulation of lightning effects on aerospace systems.

Discussion - Natural phenomena represent extremely powerful forces. The total energy of winds and tides as well as that of atmospheric electricity, in the aggregate, make even the force of a nuclear detonation seem insignificant.

The electrification phenomenon associated with lightning has been studied in varying degrees for many years, dating even beyond the early work of Franklin in 1749. At present, however, there are about as many theories as to the basic charge separation mechanism responsible for thunderstorm electrification as there are scientists specializing in the field. While much also remains to be learned about the discharge mechanisms, there is reasonable agreement about the lightning discharge channel formation and its structure. The major portion of past studies of the effects of lightning has been concerned with far-zone sferics and direct strokes to transmission lines. For the past 15 years physical and electrical effects on aircraft have also been studied in reasonable detail. It has only been in the past two years, however, that low-level lightning effects on aerospace systems have been given serious consideration. As a part of their work, the Air Force has observed that there are very little, if any, data available in two areas of lightning pertinent to its affect on aerospace systems - - close-in data for ground strokes, and adequate data on the complete spectra of lightning from DC to the low megacycle range.

Without dwelling at all on the mechanism through which charge separation occurs, it might be well to briefly discuss the nature

of lightning. As a thundercloud develops, the electrical potential difference between the cloud base and the ground continually rises to as much as 10^7 to 10^8 volts. Thus impelled, electrons begin to leave the cloud; they travel through the air and reach the ground through innumerable, mostly invisible, point discharges. But these may be insufficient to reduce the rapidly growing voltage gap. When the electric pressure gets too great, the air resistance is completely overcome and a torrent of electrons - - lightning - - passes to the receptive earth below. This occurs when a potential gradient of between 15 and 30 KV/cm causes dielectric breakdown of the air associated with these conditions. The average thunderstorm fires anywhere from 10 to 20 strokes of lightning per second. Each discharge lasts less than half a second and releases greater than 6 x 10^{16} ergs of energy. This amounts to a total release of electrical energy, from a typical thundercell, of perhaps to 10^{18} ergs per second (or 10^8 kilowatts), which can result in peak current greater than 300,000 amperes from a single lightning stroke and a total charge transfer of perhaps as great as 1000 coulombs.

A lightning discharge may be considered in terms of the typical effects produced by its various components; the intense electric field and streamering produced by the high voltages, the magnetic forces and field of the high currents, the heating and blast effects of the high-energy intermediate currents, and the metal erosion of the long-duration lower currents which continue for periods of up to one second following the short-duration high-energy components - - these are the forcing functions affecting aerospace systems.

To properly evaluate and understand the effect of the natural phenomenon of lightning on aerospace systems, several programs have been initiated and conducted during the past two years.

At the outset of the Air Force's lightning investigation, it became immediately clear that there was a complete lack of information that could be used in analysis. The power companies had been concerned with protecting and maintaining reasonably consistent operation of high-capacity power circuits. Whereas the utilities can withstand a multikilovolt transient on a power line, the Air Force is trying to ensure the protection of 30,000 erg squibs, and cannot tolerate a transient of greater than a few millivolts. Even the telephone company faces a different problem than does the Air Force. They are willing to accept some outages or to switch to alternate circuits. The Air Force goal is 100 percent reliability, without any downtime for aerospace systems.

The initial Air Force effort was associated directly with the Jupiter weapon system in its overseas environment, as shown in Figure 1. The program had two major goals. First was a fundamental program to obtain atmospheric physics data to enable a better understanding of the problem faced. The second goal was to obtain quantitative measurements of the induced effect within the Jupiter weapon system when it was subjected to near or direct lightning strokes. However, an unfortunate pattern was set by this initial effort. During the eight months that the instrumentation system

was active, the site did not receive a single lightning strike of any significance.

Even while the overseas instrumentation installation was attempting to obtain natural data, a joint test program was initiated with the Sandia Corporation, utilizing the Jupiter training complex at Redstone Arsenal, Huntsville, Alabama, during the first quarter of 1963. This effort had as its objectives the answers to these four questions:

(a) What is the principal mode by which energy is coupled into the system?

(b) What is the mechanism by which damage occurs?

(c) What measures are most effective in providing protection for the system?

(d) Can realistic data be obtained from an abbreviated system?

Simulation will be discussed later in this paper, but for these particular tests, a damped sinusoidal current was believed to offer much more information for diagnostic purposes than a single damped pulse and thus the sinusoidal current used for the Huntsville tests. It was also believed that first-order effects and the fundamental damage mechanisms could be determined by duplicating the high peak-current portion of a "true" lightning signature.

Two generators were used to provide the wide range of currents required in these tests. Both were Marx-type generators in which a capacitor bank is charged in parallel and discharged in series. Initial low-voltage tests were conducted using the Army Signal Corps Artificial Atmospheric Generator (AAG). Figure 2.

This unit provided up to 1-1/8 megavolts at up to 12 KA. The second current generator was a stationary unit built on site by the Lightning and Transients Research Institute (LTRI). Figure 3. This larger generator provided up to 5 MV and 50 KA. Both direct-strike and indirect-strike tests were conducted.

Typical data from the tests are shown in figures 4 and 5. The objectives of the program were achieved within the limitations of the test and the answers to the four questions stated in the objectives are as follows:

(a) There appear to be at least three different modes by which lightning can affect the system:

(1) By field penetration. This was illustrated by tests which indicated the necessity for maintaining a tight continuous door joints on the missile system.

(2) By electromagnetic coupling. The loop tests (indirect strike) illustrate how system cables can, through formation of loops, magnetically couple energy into the system.

(3) By direct coupling. This was illustrated by the high-voltage breakdowns noted during discharge of the generator and are presumed to be either IR voltage drops exceeding the spacing or the charging of a stray capacity to the flashover potential.

All three modes affected the results to some degree but the predominant one appeared to be electromagnetic coupling.

(b) The initiating mechanism of damage appears to be voltage breakdown at low energies with resulting follow-currents providing the energy for gross damage.

(c) The single most effective measure taken to provide protection of a system from lightning effects appears to be that which would also be taken to provide EMR protection, namely, to enclose the system within a shielding network approaching a Faraday cage. In addition to high conductivity, the shielding should be capable of conducting high currents. The test results also suggested that multiple grounds can, by reducing the direct strike current level flowing in any given cabling, reduce the severity of effects on the associated ground equipment connected to that cabling. If the missile skin integrity is not carefully maintained, the result of missile grounding may produce damaging effects even at low direct-strike current levels. To minimize these effects, joints, openings, and other discontinuties in the skin should be held to a minimum. The conductivity of joints should be as high as possible, i.e., anodizing of surfaces should be discouraged. Disconnecting all cabling to the missile appears to be the most effective protective measure that can be taken to preclude damage to the circuitry internal to the missile. Of course the associated ground equipment would still be subject to lightning effects. The indirect (loop) tests, to simulate effects produced by the radiated energy from a strike, suggest that canopies and other "diverter" schemes by themselves will not be totally effective in preventing system damage.

(d) Realistic data on the lightning response of the circuitry internal to the missile appear to be reproducible only with a total system, viz., eliminating cabling to the missile reduced the lightning

signals within the missile to an almost undetectable level. However, once the signal levels in an internal circuit are determined should additional analysis be desired it may be possible to conduct tests on that part of the missile system using the detected signal level for an input to reproduce the effects on that system.

Although field reports of incidents involving the SM-78 would suggest that direct lightning strikes are necessary to produce damaging effects in the nuclear warhead, the Huntsville test series indicated that indirect strikes may produce similar effects. The Huntsville tests also emphasized the usefulness of a transportable high-current generator in conducting tests on a system to indicate potential weak points. The tests also pointed out the value of transportable generating machines in assuring the effectiveness of measures taken for lightning protection.

The Huntsville work cannot be applied directly and indiscriminately to other missile-system evaluations. The work did, however, point the way to future tests. The Huntsville tests were never considered to be more than an initial step in promoting a closer understanding of lightning and its effect on missiles and nuclear ordnance. All the initial questions had achieved at least general answers. Lack of available time precluded the full achievement of useful data at the Huntsville Jupiter training site. The tests did establish a guide for additional effort in two major areas - the desirability of utilizing a transportable simulator and the desirability of establishing a missile test bed. The simulator will be discussed

later. Briefly, the AFWL missile test bed will now be described.

The limited amount of work conducted at the Jupiter Training Site proved most beneficial. The combination of this fair success coupled with the coincident Jupiter operational phase-out allowed AFWL to obtain a complete Jupiter installation (Figure 6) for further testing. This "missile test bed" is now nearing "operational" status (Figure 7) on its Sandia Base site. It will be utilized to obtain generic missile data on not only lightning effects, but also EMR and EMP. Although based on an obsolete system, it contains the elements of a typical missile system and will provide a useful in-house tool.

To fill-in the earlier comment about having set an unfortunate pattern with our initial Jupiter instrumentation project, two associated programs should be briefly mentioned. First, as a further adjunct to gathering natural data, AFWL has established a buried cable loop in an area likely to be struck by lightning. This is located at Socorro, New Mexico, on the campus of the New Mexico Institute of Mining and Technology (NMIMT), figure 8. The cable loop is laid out and instrumented, (Figure 9) to obtain both lightning and EMP correlation data. Through a cooperative contract, physics data on any strike is provided by NMIMT. Now in its second season of exposure, the loop has yet to see even a near stroke.

The second program is our participation in Project ROUGH RIDER. This is an ASD/AFCRL program associated with storm formation, theory, and turbulence. Our interest is associated with the thunderstorm penetration by an F-100F. The goal is to have the F-100 actually trigger or intercept

lightning strokes. In past years, the aircraft has been struck numerous times. However, now that we, in conjunction with the Sandia Corporation, have added active lightning instrumentation to the program, the aircraft has not been struck. The aircraft is shown during ground checkout in Figure 10. We are beginning to believe that the surest lightning protection is "instrumentation by AFWL"!

The total real-time simulation of a severe natural lightning stroke is an extremely difficult task. A company with which AFWL has worked closely the past two years, Lightning and Transient Research Institute, has developed a reasonable overall simulation. Through some 20 years of experience, they have evolved superposed multiple components to reproduce the magnetic forces, blast effects, and metal erosion of natural lightning discharges. This is shown in figure 11. Present military specifications for aircraft lightning arrestors spell out this as test criteria for physical integrity of protective schemes. Recent evidence, obtained by LTRI and others, however, has indicated natural strokes of much greater magnitude. Thus, where serious hazards could result from the failure of a lightning protection system, there is perhaps a need for higher-energy simulation facilities. However, the simulation techniques used depends on the particular effect of concern.

Since the Air Force's major interest is associated with the surges induced by the high rates of rise of the lightning transient, AFWL has had developed a 10-megavolt, 100-kiloamp transportable surge generator. This unit, designed and fabricated by LTRI under a joint

ASD/BSD/AFWL program, is shown in figure 12 and 13. This generator, a Marx machine, was delivered in June, 1964. The design of it is such as to provide for considerable flexibility as to the voltage, current, and ringing frequency. By utilizing various connections, the frequency can be varied from 1 KC to 1 MC - all at a peak current of about 100 kiloamps. Should it be required, the generator can be operated in a series/parallel arrangement to provide surge currents of up to 500 kiloamps. The generator has a maximum energy storage of 420 kilowatt seconds. It can be utilized in many different ways, direct, firing into loops, wires, etc. (Figure 14), to create the particular effect of interest. Combination of these techniques will be used to ascertain many aspects of aerospace system vulnerability. Missiles will be evaluated in-silo, on the surface, and even in simulated flight not only for lightning effects, but also for EMR and EMP effects.

As noted earlier, one facet of information that is lacking in the total understanding of lightning is a complete power spectral density determination close-in to severe lightning discharges. It has been clearly established that the major portion of the energy radiated from lightning is contained below 100 KC. However, to fully evaluation susceptibility of a weapon system, all energy contributions must be considered. Even though for frequencies above one megacycle, for example, the energy content is small, the coupling efficiency into a missile system may be large. Thus, even a seemingly insignificant radiation in the low megacycle range could have deleterious effects while the more energetic components at lower frequencies might not.

It is to the filling of these gaps as well as the further understanding of coupling modes that future efforts in AFWL lightning analysis will be directed. The capabilities of our powerful new simulator will be fully understood only as the machine is applied to specific problems. Plans are now being completed for extensive testing on our ballistic missile systems utilizing the AFWL surge generator and other simulation techniques. Also, we should eventually achieve the cooperation of the elements and obtain significant natural data to enable us to draw a complete, sharp picture of one of nature's most powerful expressions. All of these efforts will eventually allow the Air Force to better understand the effect of lightning on aerospace systems. With this understanding, we will be able to take the necessary steps to achieve our goal - 100 percent protection 100 percent of the time. Until that goal is achieved, nature will still plague us with - Lightning - An Aerospace Nemesis!

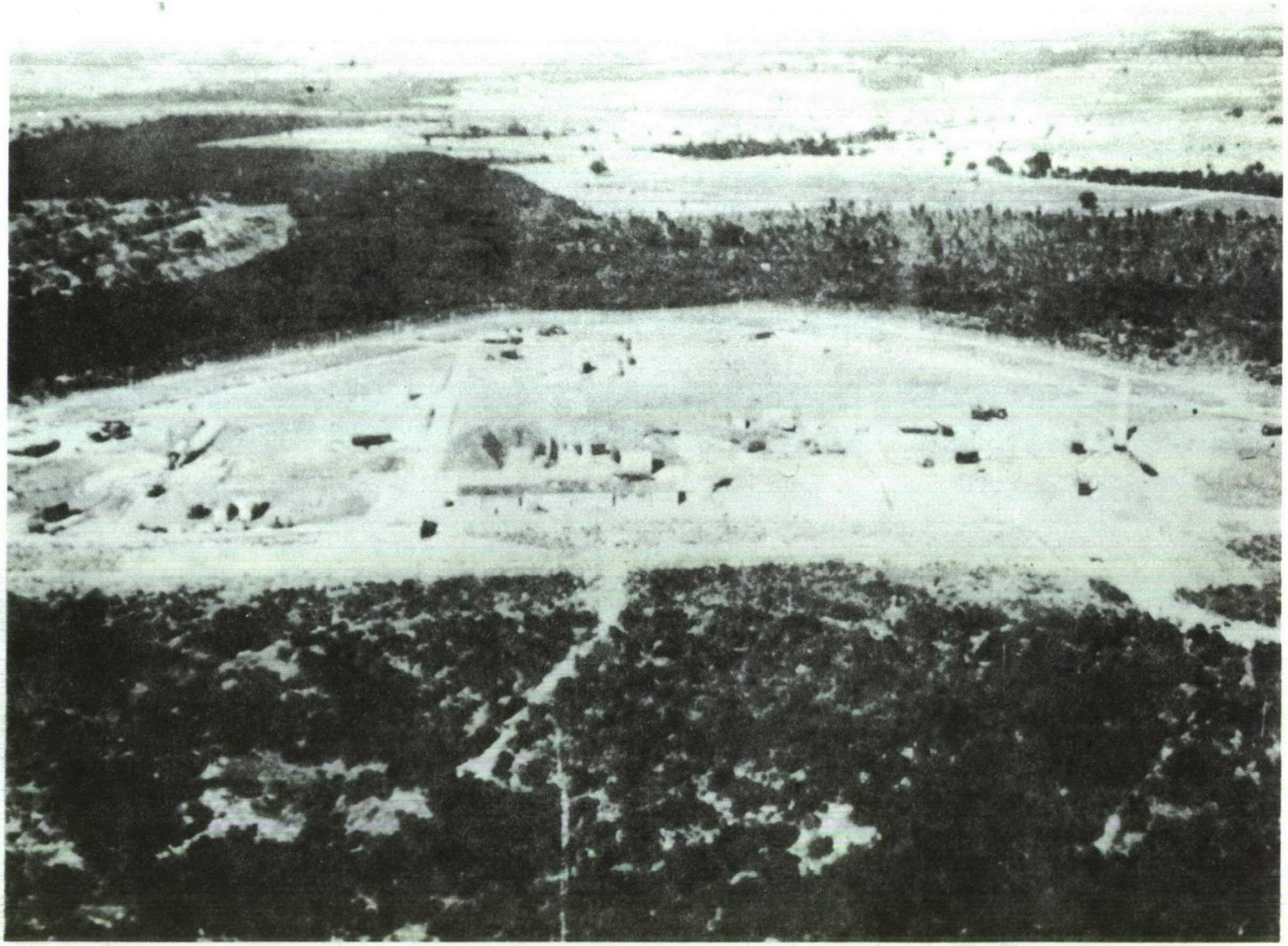


Figure 1. Overseas Jupiter Site

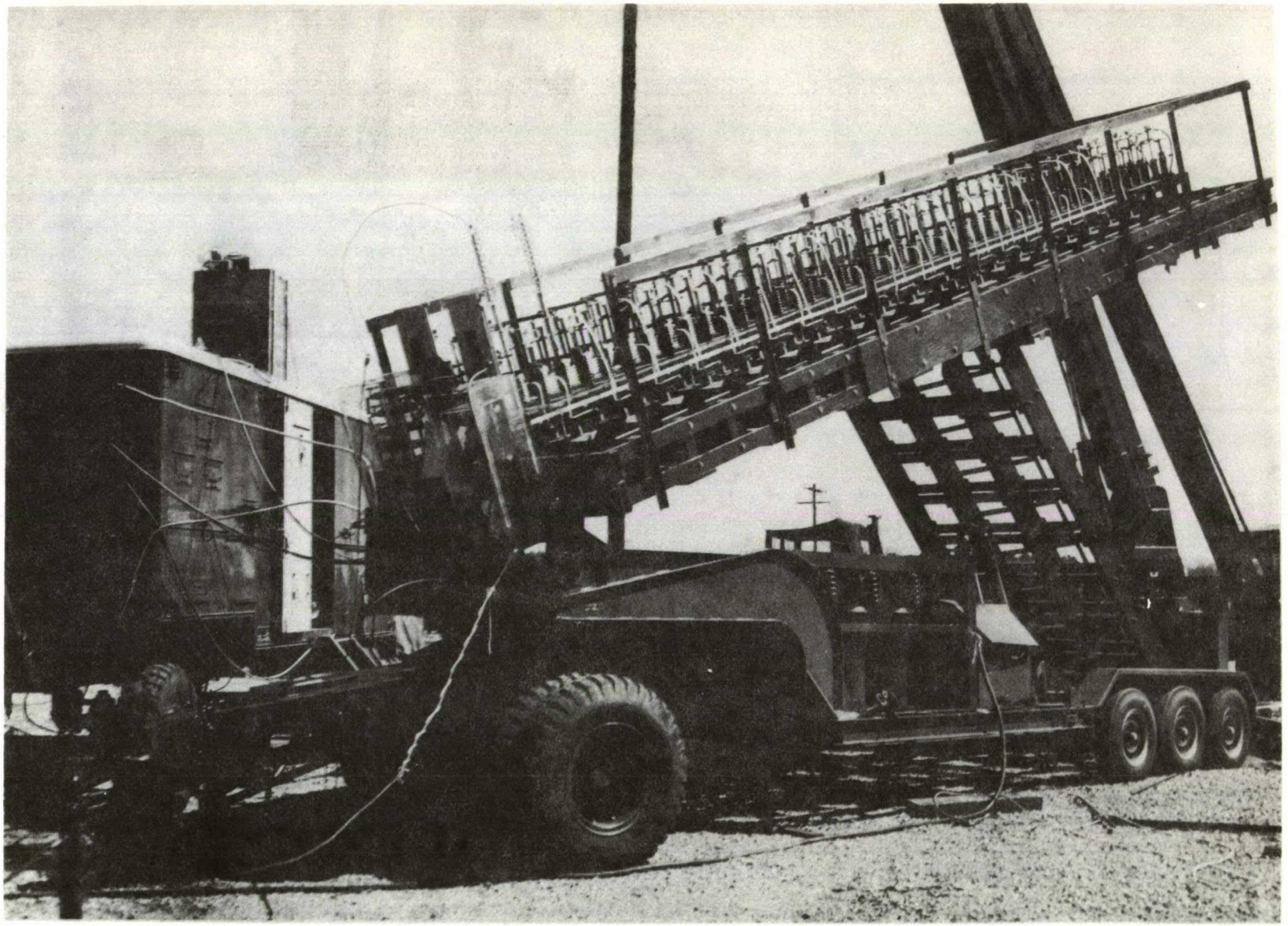


Figure 2. 1.125 megavolt Artificial Atmospheric Generator

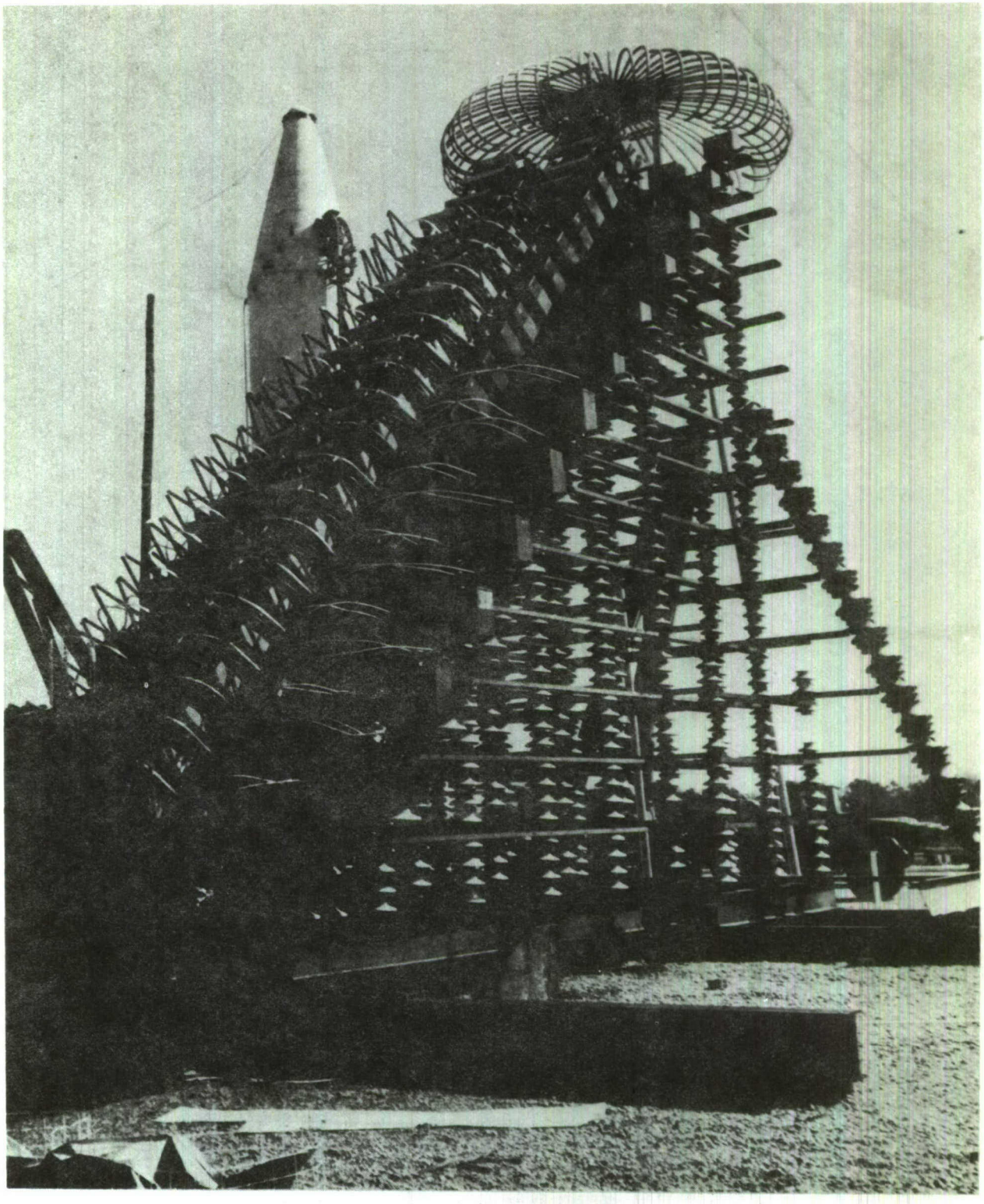
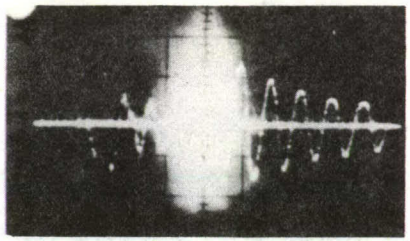
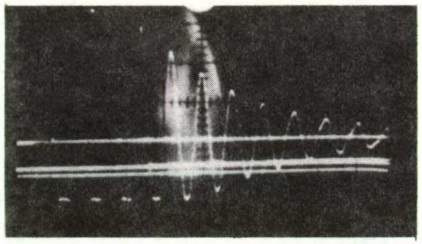


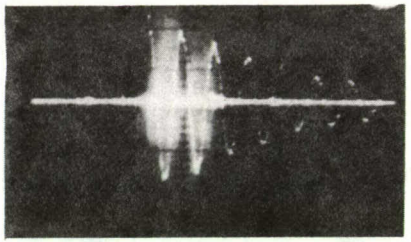
Figure 3. LTRI 5 Megavolt Surge Generator



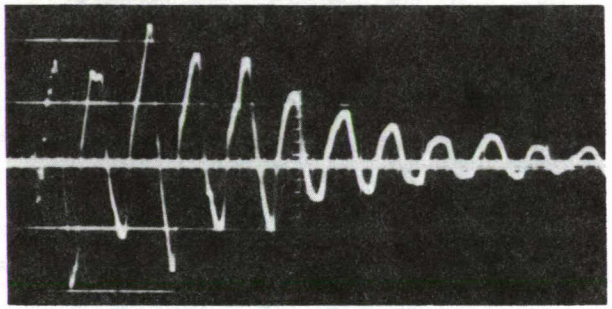
A - CONTINUITY LOOP I
200 MA/DIV - 10 μ SEC/DIV



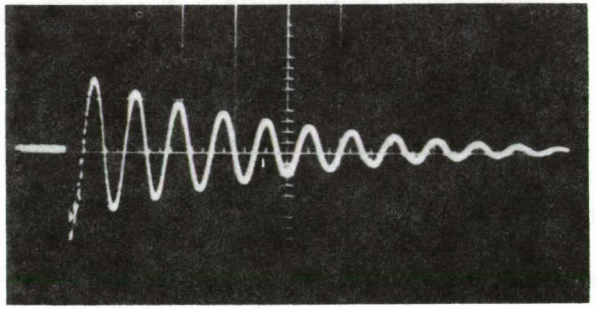
B - CONTINUITY LOOP VOLTAGE
TO GROUND
500 V/DIV - 10μ SEC / DIV



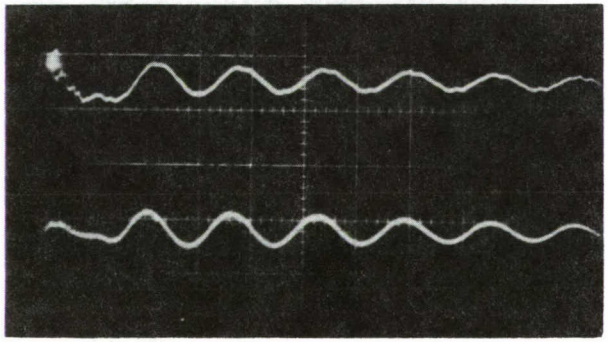
C - SQUIB I - WARHEAD
20 MA / DIV - 10μ SEC/DIV



D - D_4 BUS VOLTAGE - LCT
50 V/DIV - 10μ SEC/DIV



E - GENERATOR OUTPUT
8400 A/DIV-10 μ SEC I DIV

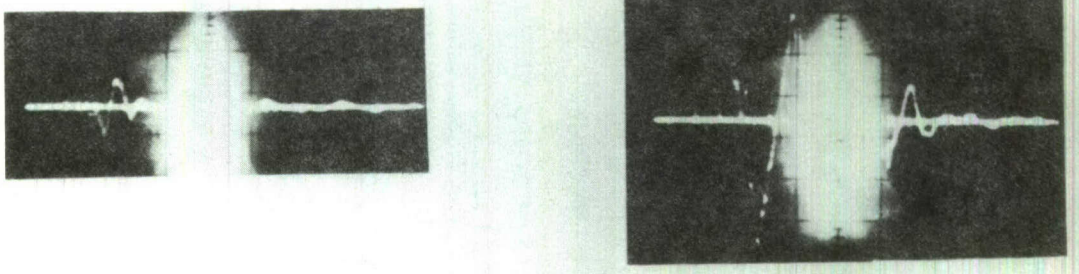


F - TOP - PHONE LINE - 2000 V/DIV
BOTTOM - AC LINE - 500 V/DIV
5 μ SEC / DIV

FIGURE 4 TEST 108
INPUT TO BIRD - OUTPUT AT BASE
1,250 KV - 1/2 CAPACITORS

(1) LOOP

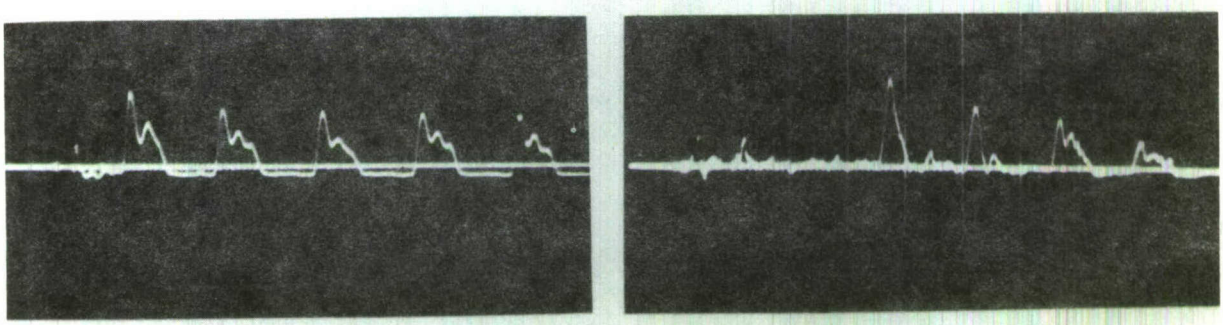
(2) DIRECT



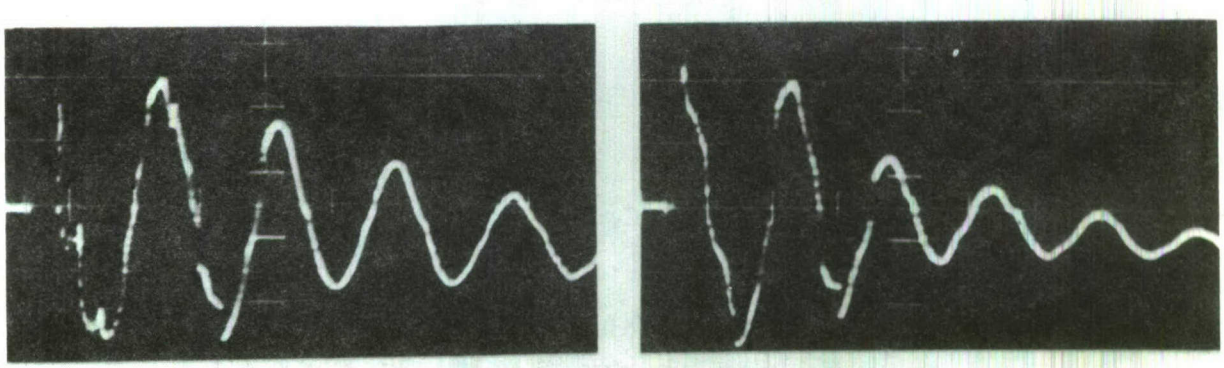
A - CONTINUITY MONITOR LOOP - 1A/DIV



B - PRESSURE MONITOR LOOP - 1000 V/DIV



C - VOLTAGE ACROSS POWER DIODE - HPT - 200 V/DIV



D - GENERATOR OUTPUT - 24 NA / DIV

FIGURE 5 TESTS 144 (NW) & 145
COMPARISON OF LOOP VS DIRECT FEED
2.1 MV, 50 KA
10 μ SEC / DIV

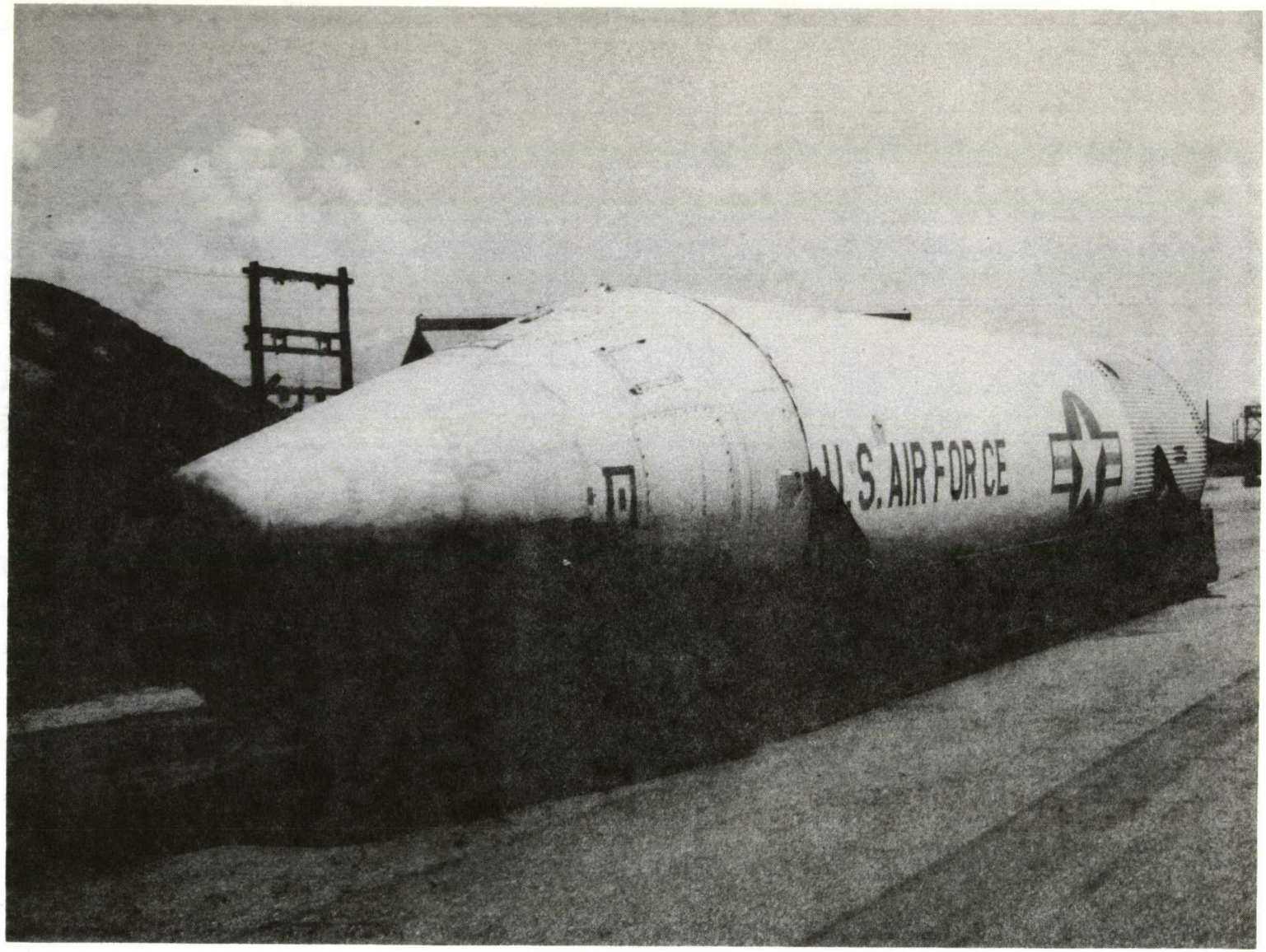


Figure 6. AFWL Jupiter Missile

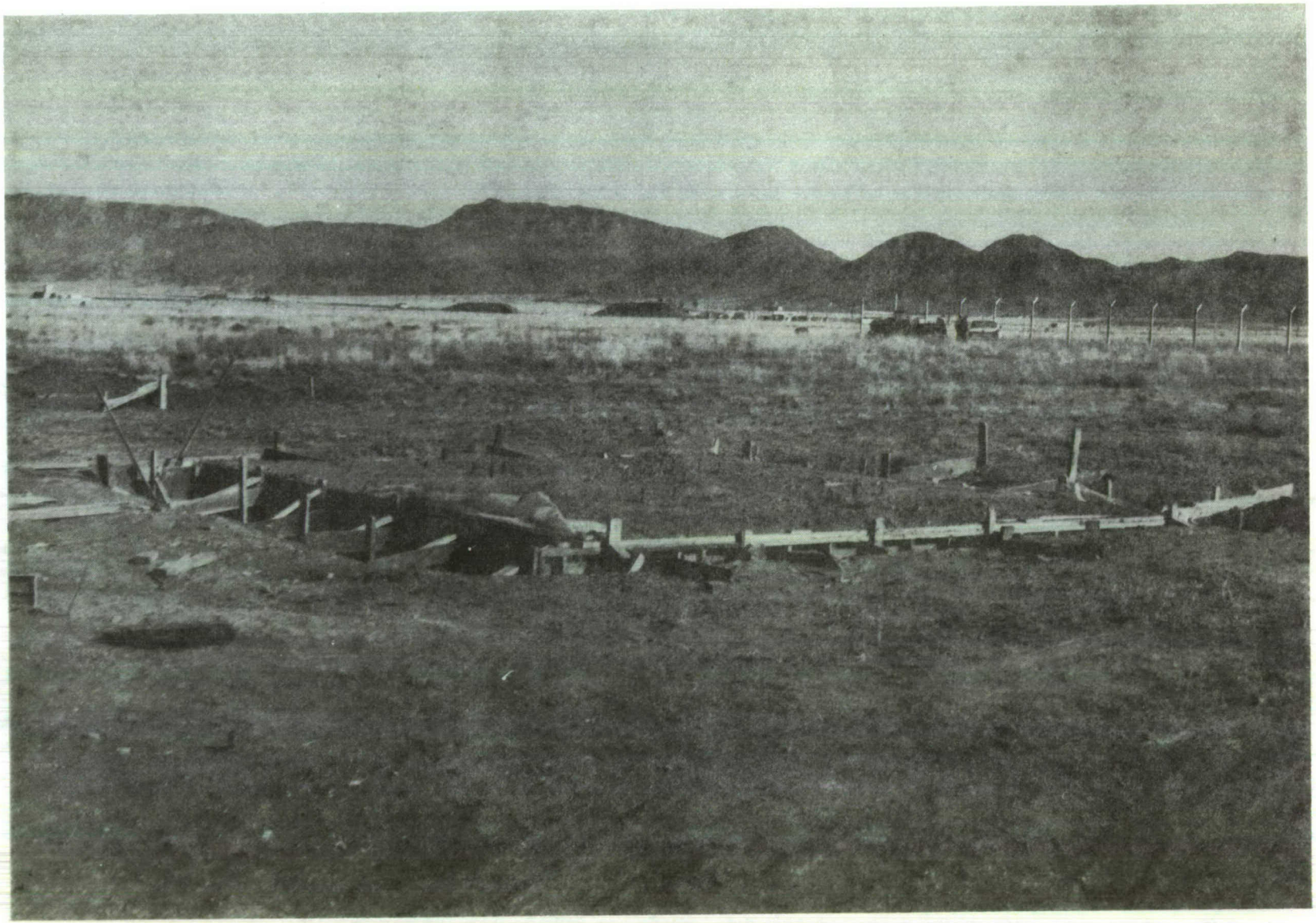


Figure 7. AFWL Missile Test Bed



Figure 8. AFWL Cable Site, NMIMT

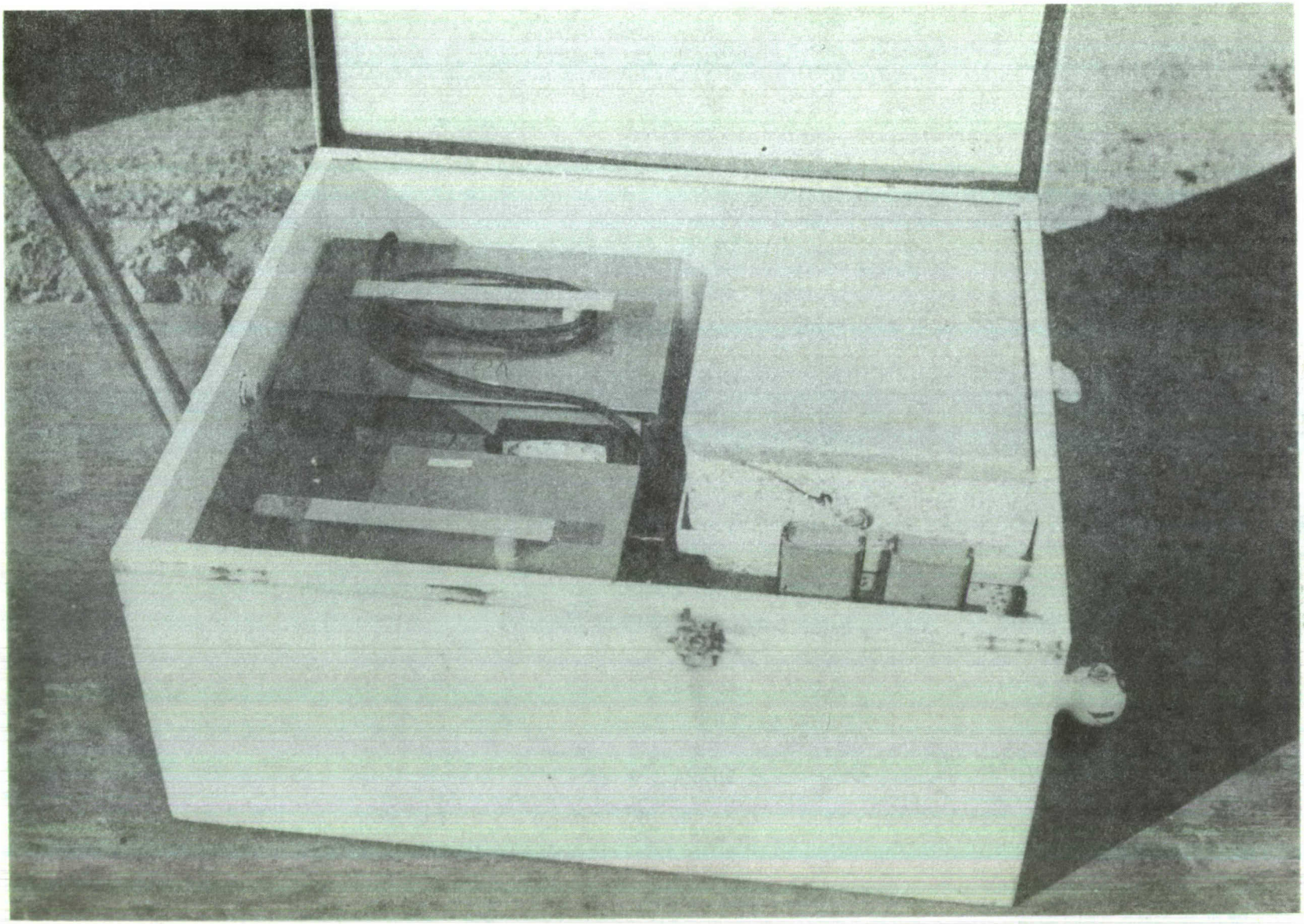


Figure 9 AFWL Cable Loop Instrumentation (NMIMT)

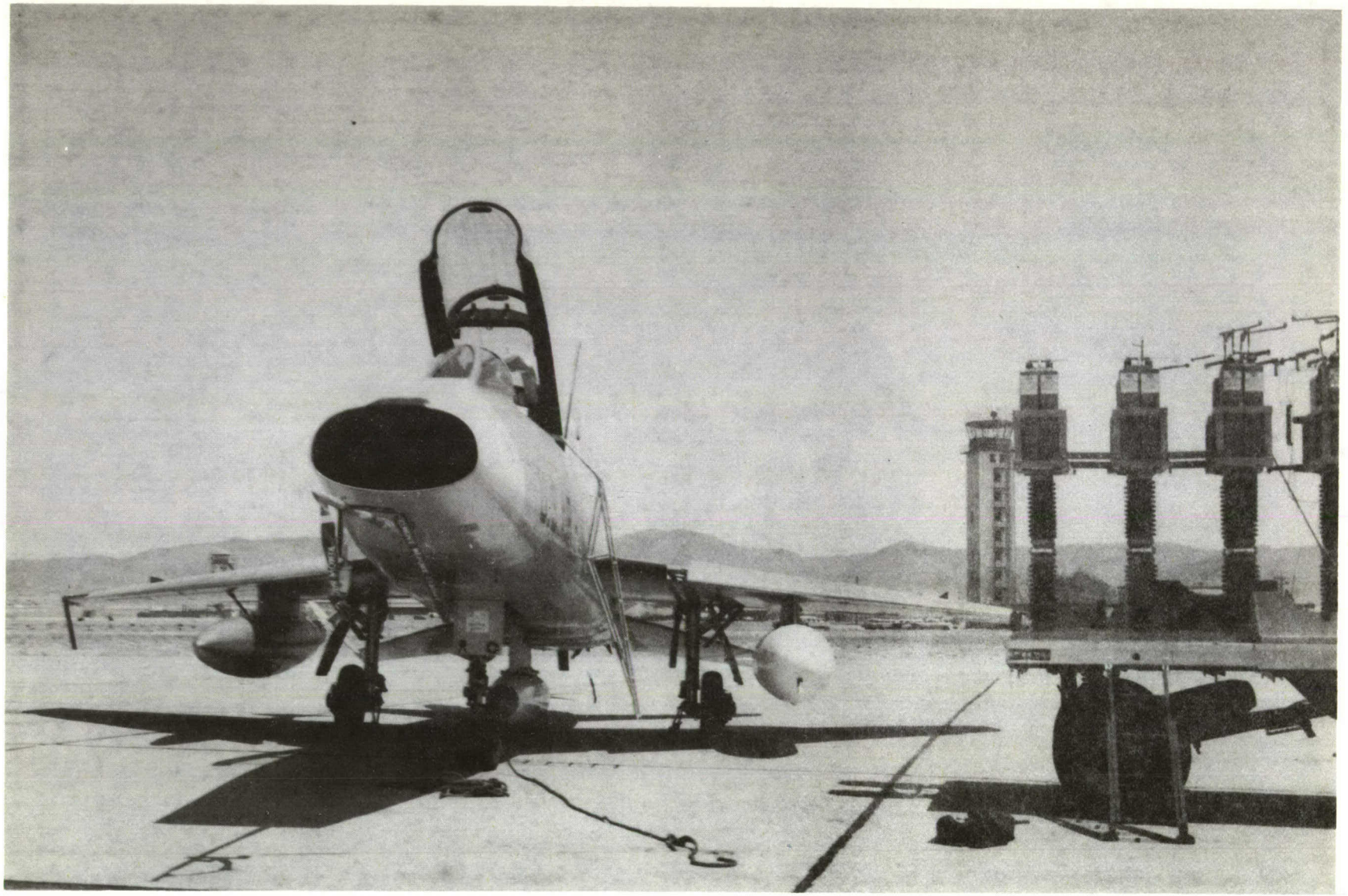


Figure 10. Instrumented ROUGHRIDER F-100F Aircraft

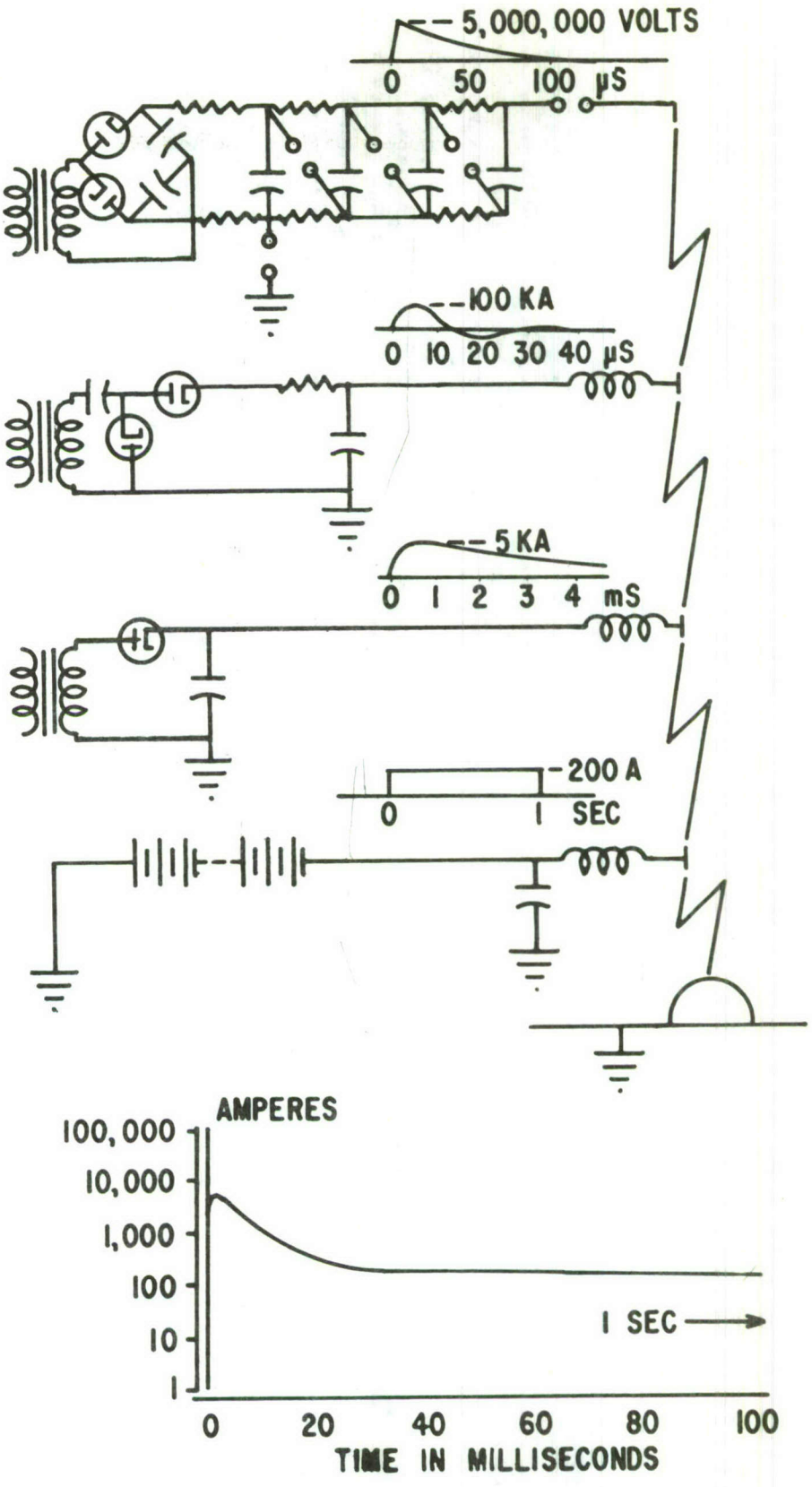


Figure 11 Artificial Lightning Generator with Multiple Current Component Sections and Resultant Composite Output Current Waveform

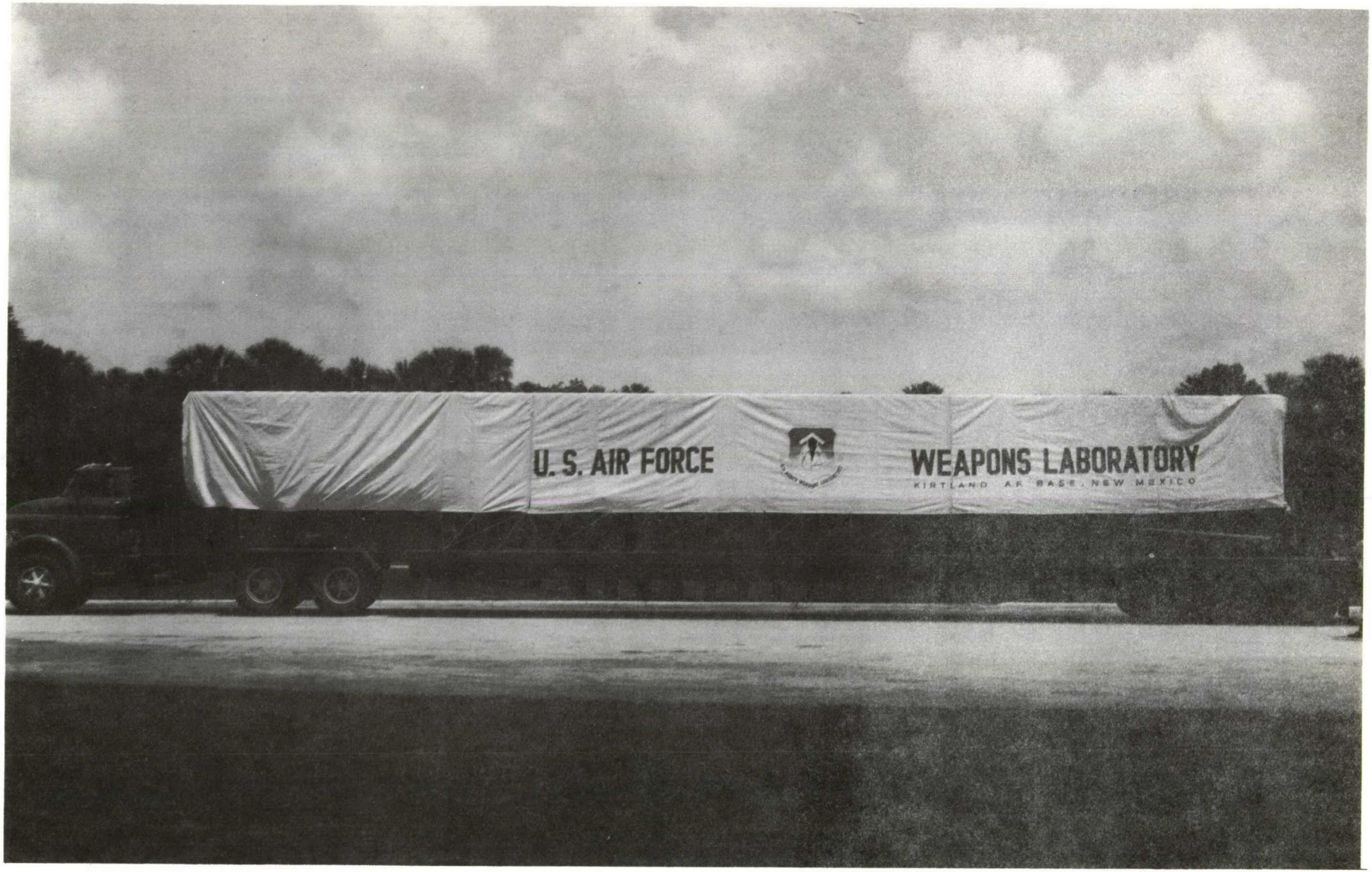


Figure 12. AFWL Surge Generator

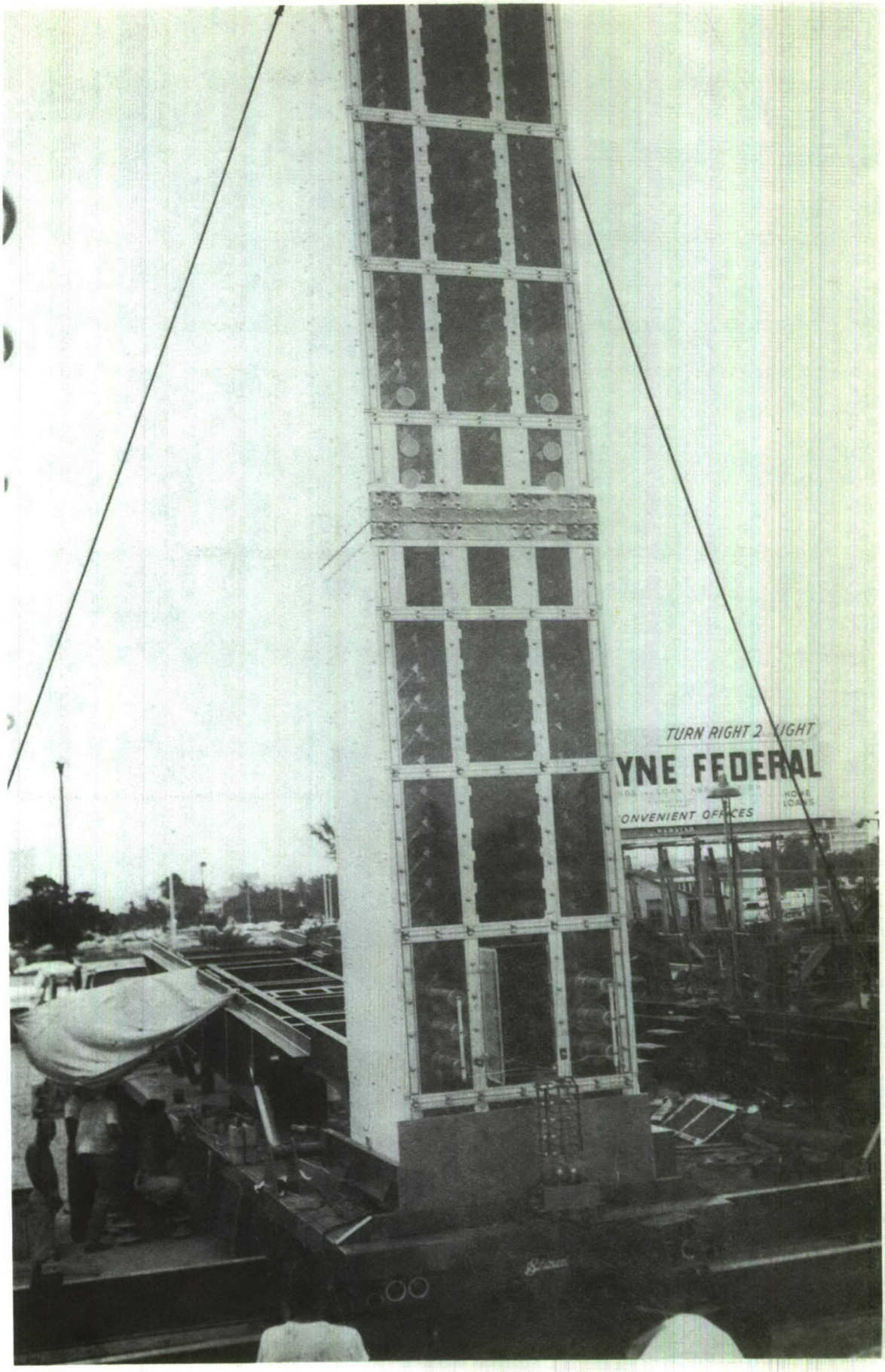


Figure 13. AFWL Surge Generator

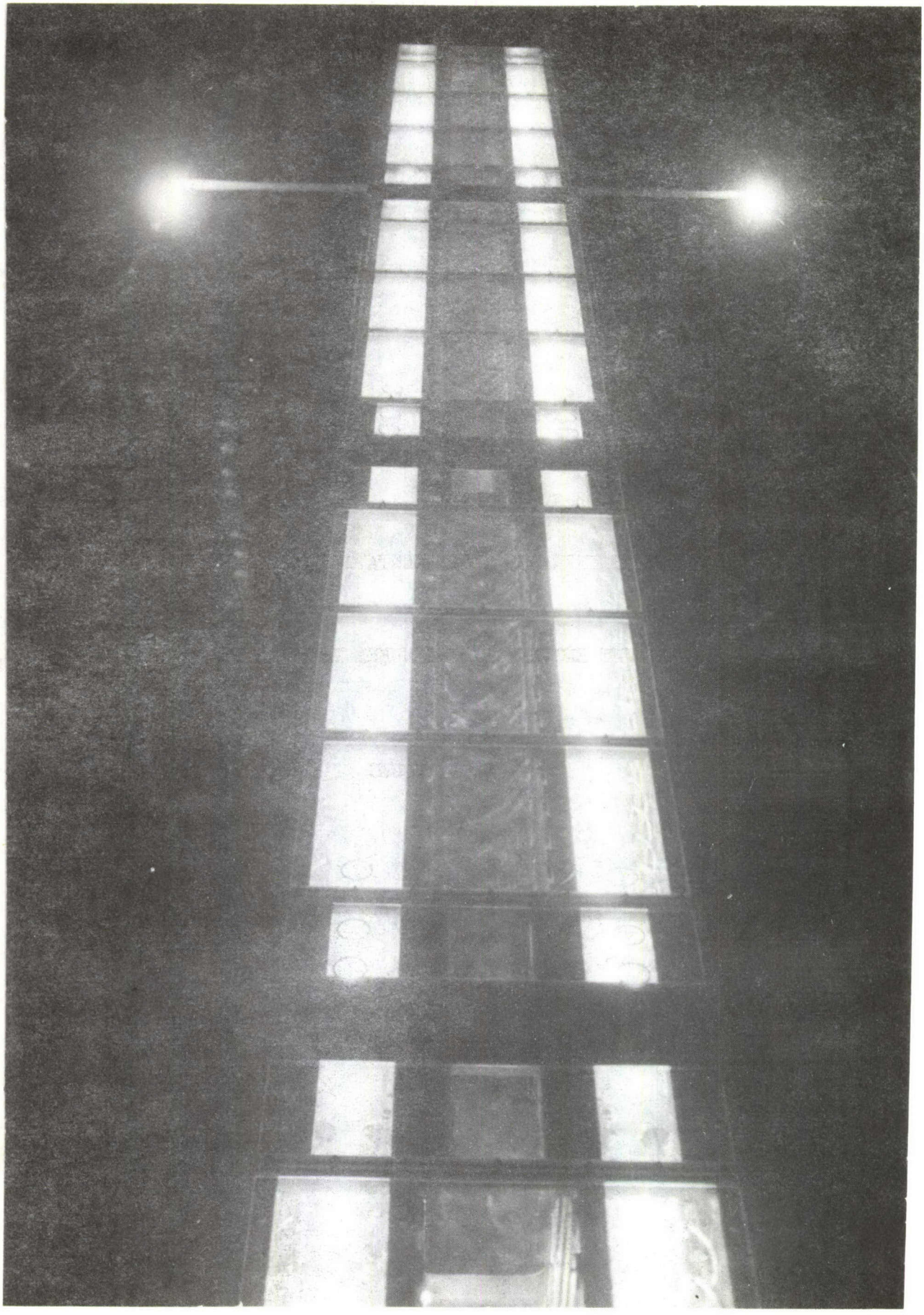


Figure 14. Actual Loop Firing, AFWL Surge Generator

USE OF INFRARED INSTRUMENTATION FOR DETECTION OF JET ENGINE TURBINE WHEEL TEMPERATURE

F. L. YOUNG

NOMENCLATURE

A = radiating area

A_1 = outside area of tailcone

A_2 = outside area of pyrometer housing

a = Stefan-Boltzman constant

C = Configuration factor to allow for relative position and geometry of bodies

C_p = specific heat of stainless steel

C_{pi} = specific heat of insulation

C_{ps} = specific heat of steel

C_{t_1} = thermal capacitance of tailcone

C_{t_2} = thermal capacitance of steel housing

C_{t_3} = thermal capacitance of insulation

E = radiated energy

e = emittance of radiating surface

e_h = emissivity of the pyrometer housing

emf = electromotive force generated by thermopile circuit

e_t = emissivity of tailcone

Gr = Grashof number

h_{rad} = radiant heat transfer coefficient

K_s = thermal conductivity of steel

k_i = thermal conductivity of insulation

L = length of pyrometer housing

M_i = Mass of insulation

M_s = Mass of steel housing

M_t = Mass of tailcone

Nu = Average of Nusselt number

Pr = Prandtl number

R_2 = resistance of thermopile

R_3 = resistance of nickel compensating coil

R_{t_1} = thermal resistance to radiant heat transfer between tailcone and outside

R_{t_2} = thermal resistance to convection heat transfer between tailcone and outside

R_{t_3} = thermal resistance to radiant heat transfer between the pyrometer housing and inner cone surfaces

R_{t_4} = thermal resistance to convection heat transfer between pyrometer housing and inner cone environment

R_{t_5} = thermal resistance to heat conduction through steel housing

r_1 = inside radius of insulation

r_2 = inside radius of the steel housing

r_3 = outside radius of the steel housing

S = Pyrometer output signal

T = absolute temperature

T_H = average absolute temperature of pyrometer housing

T_h = initial absolute temperature housing

T_I = average absolute temperature of inner cone surface

T_i = initial absolute temperature of the inner cone surface

T_s = absolute temperature of outside air

T_t = average absolute temperature of tailcone

T_{t_6} = thermal resistance to heat conduction through insulation

t_e = average tailcone temperature

t_f = average film temperature

t_s = temperature of outside air

Introduction

The Air Force is very much concerned about the loss of B-47 aircraft caused by the failure of the J-47 engine turbine wheel. Figure 1 is a graphic illustration of these failures. The most damaging of these occurs when a large chunk becomes separated from the wheel. This type of failure generally results in severe damage to the aircraft or its destruction. These failures are caused by a loss of cooling air to the wheel which permits the wheel to exceed its temperature limit. A brief description of the engine will provide a better understanding of how this can occur. The J-47 engine has an axial flow compressor, single stage turbine and a fixed area exhaust nozzle. Maximum thrust is delivered at a fixed RPM and exhaust gas temperature (EGT) of 1274°F. The turbine wheel temperature during operation is between 400 and 850°F. It is heated by conduction from the turbine blades located in the gas stream (Fig 2) and cooled by convection, using compressor air flowing over the faces.

The compressor is subject to stalls which reduces the mass flow through the compressor and the cooling air to the turbine wheel. When a compressor stall persists in excess of 2.5 minutes, the temperature is increased sufficiently to allow the applied stress to exceed the tensile strength of the wheel material (Fig 3); wheel failure then occurs.

Several methods have been explored to prevent this: (1) Improvement in the engine design to improve stall tendencies; this was considered prohibitive from an economical standpoint since it would require major changes to the compressor. (2) Fusible plugs and temperature sensitivity paints; the plugs and paint were inspected after each flight for indications of overtemperature. They were unreliable and did not clearly exhibit a change.

(3) Thermocouples fused to the wheel; the thermocouples output was input to a transformer mounted on the turbine shaft. The signal was transmitted to a stationary induction transformer which triggered warning lights in the cockpit. The device was inaccurate and lacked system reliability. A similar system was also tested using the high-speed slip ring technique, however, this system was so complex that it was eliminated from consideration except on short duration static tests.

The experience gained from methods of measuring turbine wheel temperatures using direct techniques dictated consideration of different approaches. These approaches were: (1) Measuring the turbine wheel temperature without direct contact and (2) Correlating turbine wheel temperature with compressor stall conditions. The use of approach number 1 would require a sensing element to be located in the exhaust cone to provide the pilot an indication of when the turbine wheel temperature was reaching a critical level. Approach number 2 required the correlation of EGT, RPM with turbine wheel temperatures during compressor stall. The first approach was ruled out for fleet application from a cost standpoint because of modifications to the engine exhaust cone, aircraft wiring and instrumentation requirements wherein the stall warning technique would use existing EGT and RPM sensors. However, an inexpensive temperature sensing technique was required to determine turbine wheel temperatures during stall since the engine test encompassed endurance items and the thermocouple and high-speed slip ring system does not lend itself to long term operation.

It was decided to explore the use of the infrared radiation pyrometer. This paper discusses the problems encountered in adapting a radiation pyrometer to detect and measure turbine wheel temperatures during this test.

Instrumentation Requirements

The detailed approach to the problem of obtaining a radiation pyrometer for sensing turbine wheel temperatures can best be discussed by considering the requirements the device must fulfill:

1. The device must operate in an environmental temperature of 1000°F and operate for a minimum of 400 hours without failure.
2. Must be accurate within ± 10°F of the turbine wheel temperature.
3. It must not interfere with existing instrumentation on the engine.

A survey was conducted to determine the available equipment for this application. The equipment that existed was considered too expensive and would have to be adapted to this application. Because of the outcome of the equipment survey it was decided to use in-house development supplemented by a contract with Oklahoma State University, Stillwater, Oklahoma.

Review of literature on instruments available in this field disclosed that the Minneapolis-Honeywell Regulator Company manufactured a commercial radiation pyrometer, under the trade name of Brown Radiamatic. The description of the instrument is as follows:

1. Model Number: Miniature Radiometer No. 933A3.
2. Over-all Dimensions: 2 11/16 inches long, 2 3/8 inches diameter.
3. Minimum target diameter at 6 inches focal length: ½ inch.
4. Normal temperature range: 500° - 3200°F.
5. Response time to 98% temperature change: 2.0 seconds.
6. Lens: Calcium Fluoride.

This device consists of lens and thermopile which uses a nickel coil to compensate automatically for any environmental temperature to 250°F. The Brown Miniature Radiamatic was chosen to investigate the feasibility of using a radiation pyrometer for remote detection of turbine wheel temperature.

Principle of the Compensated Radiation Pyrometer

The radiation pyrometer detects the temperature of a source from its emitted radiant energy. The operation of the Brown Radiamatic, as with any total radiation pyrometer, is based on the physical law that all objects radiate energy at a rate proportional to the fourth power of their absolute temperature. Basically, this energy transport is described by the equation[1]:

$$E = eaAT^4 \qquad (1)$$

However, the equation (1) describes only the energy exchange from a body of finite temperature to an environment of zero absolute temperature. When such exchange takes place between two bodies of finite temperatures both temperatures must be introduced into the expression. Considering two parallel surfaces of the same area, but different emittances, the energy transport would be described by:[1,5]

$$E_1 \text{ to } 2 = \frac{C \quad a}{\frac{1}{e_1} + \frac{1}{e_2} - 1} (T_1^4 - T_2^4) \qquad (2)$$

Equation (2) closely approximates the total radiation pyrometer and shall be used to explain the principle of operation. The radiant energy from the source of T_1 is focused by a lens onto a thermopile of temperature T_2. Associated with the thermopile is a wire resistance, designated by R_2, as shown in Fig 4. As a typical thermocouple circuit, the thermopile represents the high temperature measuring junction. The point of connection between the thermopile leads and the lead-out terminals represents the environmental temperature junction, T_3. By measuring the emf generated because of the temperature differences between thermopile and the reference junction, the radiating source temperature T_1 can be determined. To compensate for variations of T_3 a nickel compensating coil or resistor R_3 is placed as a shunt between the lead-out wires. This is described by the equation:

$$S = \frac{emf}{1 + \frac{R_2}{R_3}} \qquad (3)$$

As the emf decreases, the nickel compensating coil resistance increases proportionally and the signal output of the pyrometer S remains constant. The maximum temperature the nickel coil would compensate for was 250°F. Therefore, the requirement for an insulation scheme existed.

Selection of Lens Material

The lens material must be carefully selected since the wave length of the radiated energy is a function of the absolute temperature of the body observed. A given lens material is not transparent to some wave length bands; partially transparent to others, and perhaps completely opaque to other wave length bands.[2] A relationship between the absolute temperature of a black body and the wave length at which maximum emission occurs is given by Wien's Distribution Law as follows[1].

$$\text{Wave length} = \frac{5216 \text{ microns}}{\text{Temperature (Degrees Rankins)}} \qquad (4)$$

Although this relationship is for a black body, it is reasonably good approximation for the point of maximum emission for a gray body.

If an infrared radiation pyrometer is used to detect temperature between 400°F and 1100°F, the associated wave lengths for maximum emission will be:

$$\text{At } 400^\circ\text{F, Wave length} = \frac{5216}{(400 + 460)} = 6.06 \text{ microns} \qquad (5)$$

$$\text{At } 1100^\circ\text{F, Wave length} = \frac{5216}{(1100 + 460)} = 3.34 \text{ microns} \qquad (6)$$

Ordinary glass does not transmit infrared energy at these wave lengths. The upper limit for glass is approximately 2.7 microns. Therefore, a special lens material was required that would transmit the energy in this range. Calcium Fluoride[3] is transparent to radiant energy from .3 to 10 microns and therefore was chosen as a suitable lens material.

Emittance Determination

The turbine wheels have a variety of surface conditions, therefore, it was

necessary that the temperature measuring system be calibrated for a constant emittance value.[4] Surface turbine wheels fall into four general categories:

Category 1 - New wheel surface - recently overhauled and freshly "Parkerized."

Category 2 - Wheel surface Parkerized and Oxidized - simulating typical operating conditions.

Category 3 - Wheel surface without preparation and unoxidized.

Category 4 - Uncoated and oxidized.

Again noting the basic relationship[1]:

$$E = \epsilon a A T^4$$

One specimen was chosen from each of the categories and the emittance was determined using this relationship. The emittance values were determined for incidence angles of 90° and 75° (0° and 15° from the perpendicular) for the four specimens at 600°F, 700°F, 800°F, and 900°F. It was of particular importance to measure the emittance rather than the erroneous radiosity (emitted energy plus reflected energy from surrounding surfaces.) This was accomplished by installing the insulation shield around the periphery of the sighting tube, thus excluding energy from the furnace wall. The insulation also prevented heating of the sighting tube which could also give erroneous results.

The measurements were made by heating the specimens to a prescribed temperature, allowing the temperature to stabilize, insert the pyrometer into the furnace and read the output while sighting on the specimen. The pyrometer output signal was measured in millivolts using a Leeds & Northrup potentiometer. Each of the specimens were drilled and instrumented for thermocouples to determine when conditions of temperature equilibrium were reached. The recorded value was

then divided by the value for black body radiation at the same temperature. The resulting ratio is the measured emittance.

The results of the emittance tests are shown in Fig 5. With the exception of specimen No. 1 at 0° deflection, the values for each of the specimens were quite uniform and indicate an average value of .975 could be used for the temperature range 700° - 900°F. Referring to Fig 6 at 800°F, an error of 10 degrees would require an emittance change from approximately .94 to 1.00. This implies that a constant emittance of .975 could be used for all turbine wheels and remain within approximately $\pm$ 10° temperature accuracy.

Environmental Determination

A system capable of performing in an environment of 1000°F was required. The pyrometer housing and insulation arrangement was designed for operation of the pyrometer below 250°F even though the external temperature in the inner cone reached 1000°F. The housing was designed to contain sufficient insulation to keep the pyrometer from exceeding 350°F during the interval after engine shutdown when the compressor cooling air ceases and the tail cone is exposed to exhaust gases of approximately 1300°F. The temperature limit was selected for two reasons. First the pyrometer would not be required to operate during this period, therefore, the compensation limit of 250°F could be exceeded. Secondly, any temperature above 350°F might damage the instrument. The insulation material selected was Armstrong A-28 Fire Brick. The temperatures, after engine shutdown, were approximated using a thermal analyzer to determine the maximum temperature the pyrometer would reach. A resistance capacitance heat transfer simulator was very useful in analyzing the transient heat transfer problem.

The first consideration was the heat transfer from the hot tail cone to the outside air. This energy transfer is by convection and radiation, and the tail cone was assumed to be a horizontal cylinder. If the free convective flow

is assumed to be laminar, the average Nusselt number can be obtained from:[1]

$$Nu = .55\ (GrPr)^{.25} \qquad (7).$$

The physical properties of the Grashof number, k of the Nusselt number, and the Prandtl number were evaluated at the average film temperature given by[1]:

$$t_f = t_s + \tfrac{1}{2}\ (t_e - t_s) \qquad (8)$$

Thus, the average heat transfer coefficient h_{conv} was obtained. To determine the amount of radiant heat transfer between the tail cone and the outside air, it was convenient to use a radiant heat transfer coefficient defined by[1]:

$$h_{rad} = e_t a\ (T_t^2 + T_s^2)\ (T_t + T_s) \qquad (9)$$

The thermal resistances to convection heat transfer and radiant heat transfer are calculated from their respective heat transfer coefficients and the outside area of the tail cone. Thus:

$$T_{t_1} = \frac{1}{h_{rad}\ A_1} \qquad T_{t_2} = \frac{1}{h_{conv}\ A_1} \qquad (10)$$

The thermal capacitance of the tail cone is calculated by multiplying the mass of the tail cone by its specific heat.

$$C_{t_1} = M_t C_p \qquad (11)$$

The second concern was with the convection and radiant heat transfer from the inner cone to the pyrometer housing and then the conduction of the heat through the housing to the pyrometer itself. For ease in calculations as well as a safety factor, the environment of the tail cone was assumed to be the same temperature as the inner cone surfaces, 1300°F. However, during engine operation the actual environment of the inner cone was 1000°F, and the surface of the pyrometer housing was assumed to be at that temperature the instant of shutdown. The convective heat transfer coefficient was then calculated from Equation (7)

with the properties evaluated at the arithmetic average of the housing surface temperature and the inner cone assumed temperature. The determination of the radiant heat transfer was more difficult in that the temperature of both surfaces varied. The equation used was:

$$h_{rad} = e_h a \, (T_1^2 + T_H^2)(T_1 + T_H) \qquad (12)$$

The average temperatures were estimated to be:

$$T_I = T_h + 3/4\,(T_i - T_h)$$

$$T_H = T_h + 1/4\,(T_i - T_h) \qquad (13)$$

The thermal resistances to convection heat transfer and radiant heat transfer between the pyrometer housing and the inner cone environment and surfaces are:

$$R_{t3} = \frac{1}{h_{rad} \quad A_2} \qquad R_{t4} = \frac{1}{h_{conv} \quad A_2} \qquad (14)$$

The thermal resistances for the heat conduction through the pyrometer housing are:

$$R_{t5} = \frac{\ln \frac{r_3}{r_2}}{2 \quad K_s L} \qquad R_{t6} = \frac{\ln \frac{r_2}{r_1}}{2K_i \quad L} \qquad (15)$$

The thermal capacitances of the components of the housing are:

$$C_{t2} = MsCps \qquad C_{t3} = M_i C_{pi} \qquad (16)$$

The entire heat transfer occurring in the system is described and must be converted into the analogous electrical circuit. The conversion was done by scale factors defined as:

$$S_R = \frac{\text{thermal resistance}}{\text{electrical resistance}} = \frac{R_t}{R_e}$$

$$S_C = \frac{\text{thermal capacitance}}{\text{electrical capacitance}} = \frac{C_t}{C_e} \qquad (17)$$

$$S_T = \frac{\text{thermal time}}{\text{electrical time}}$$

Also by the analogy:

$$S_T = S_R S_C \tag{18}$$

By assuming values of S_R and S_C and using equations (17), the analogous electrical resistances and electrical capacitances are determined. By equation (18) the scale factor linking the electrical circuit time and the thermal time is determined. The electrical circuit shown in Fig 7 was then put on the thermal analyzer. The potential in the thermal and electrical systems are also analogous resulting in:

$$S_p = \frac{\text{temperature}}{\text{volts}} \tag{19}$$

By assuming a value of S_p, the voltage produced in the electrical circuit can be linked to actual temperature

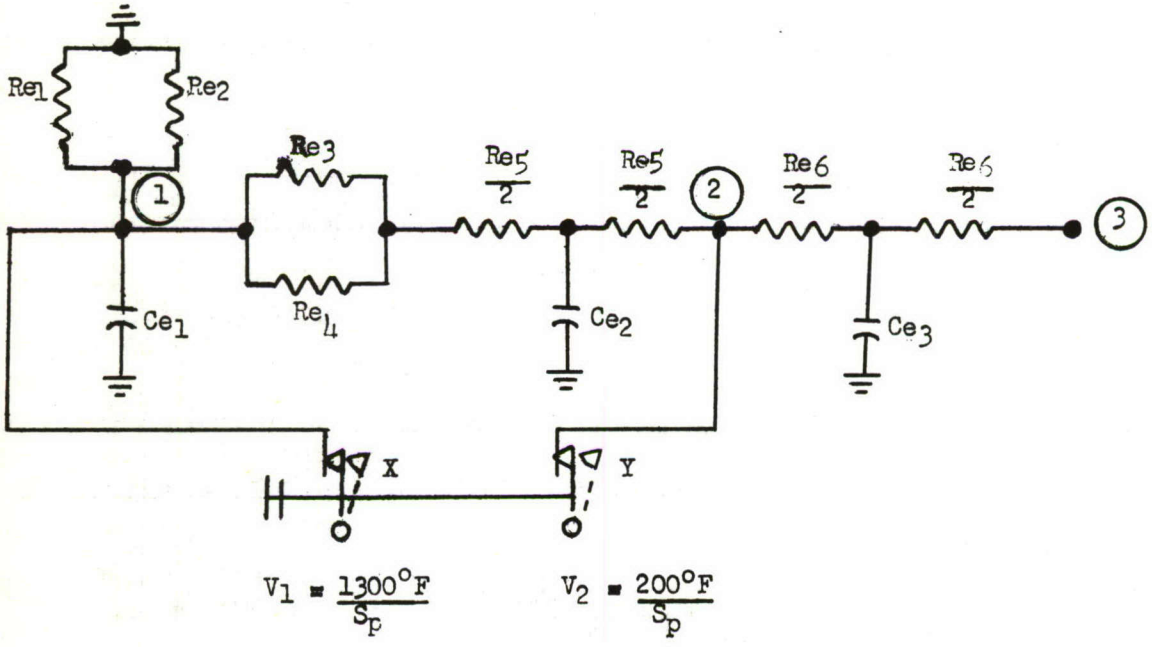


Figure 7 Thermal Analyzer Circuit

In the circuit, V_1 represents the initial temperature of the tail cone, while V_2 represents the temperature of the insulation. Up to engine shutdown, the insulation was at 200°F because of the cooling air. At shutdown, relays X and Y are broken and the circuit is allowed to come to equilibrium. The voltages at nodes 1 and 3 are recorded and presented as temperatures in Fig 8. Since the temperature limit of 350°F was not exceeded, this design was deemed acceptable, and the housing was built as shown in Figs 9 and 10.

Cooling Air Requirement

The only access area to the inner cone is through the strut. At 100% engine rpm, the metal strut surfaces reach approximately 1300°F, which causes considerable heat conduction through the insulation into the incoming cooling air.

To provide a better understanding of the temperature rise of the cooling air as it passed through the high-temperature region of the tail cone, thermocouples were placed at intervals in the ducting and on the pyrometer housing. Also, to determine the quantity of cooling air required, a square-edged orifice flow-meter was installed in the cooling air line.

The cooling air and pyrometer housing temperatures were first monitored using a dummy shell. The results of the cooling air requirement test data (Fig 11) show that 0.04 lbm/sec supplied at an inlet temperature of 150°F or below is sufficient to keep the radiation pyrometer within the 250°F temperature limit. The cooling air requirement is insignificant when compared with the total compressor air mass flow.

Another problem associated with the adaptation of a radiation pyrometer was the static pressure in the tail cone. The static pressure rises to 21 psia at 100% engine rpm (Fig 12). The difficulty was that cooling air must not only be supplied in a sufficient quantity, also a sufficient pressure to overcome the 21 psia in the inner tail cone area. By tapping the fifth stage of the compressor,

air is available at 25 psia and 150°F. Thus, there was no cooling air problem.

Engine Tests

The pyrometer was installed in the cone (Fig 13) for a preliminary test to determine if it could accurately detect turbine wheel temperatures and withstand the inner exhaust cone environment during engine operation. There was good correlation between wheel temperatures obtained from previous tests using thermocouples embedded in the wheel and pyrometer readings. The results of the preliminary tests (Fig 14) were considered satisfactory and the pyrometer was installed for the 1000 hour test in the as is configuration. The pyrometer output (Fig 15) was recorded on a Leed & Northrup Azar Recorder. The system worked perfectly for 42 hours of the test but the output signal became erratic and stopped at this point. It was removed at 150 hour engine tear down. Inspection revealed the sighting tube of the pyrometer housing had broken loose and the installation shattered which accounted for the erratic signal.

The pyrometer housing was redesigned (Fig 16) and the sighting tube welded with tungsten inert gas process to reduce possibility of breakage. The system was reinstalled and operated satisfactorily for the remainder of the test which terminated at 893 hours due to failure of an engine component. The EGT, RPM and turbine wheel temperature relationship during a compressor stall was established (Fig 17) using wheel temperatures obtained from the pyrometer. This was accomplished by intentionally stalling the engine. The data obtained during the compressor stall provided the information required to prepare a specification for a stall detection device. This device will provide the pilot a warning of stall in sufficient time to recover the engine and prevent over-temperature damage to the turbine wheel. The information also provided the time delay requirement for the stall detection device to prevent false indications of stall from transient engine operation.

Concluding Remarks

Results of test prove that total radiation pyrometer can be used in a rather severe environment to continuously detect turbine wheel temperatures. A magnetic amplifier and trigger amplifier could convert the output signal from the pyrometer to degrees Fahrenheit and trip a warning light circuit at a given temperature to provide a warning when the turbine wheel reached a set temperature limit. The pyrometer is inexpensive, however, the exhaust cone modification, aircraft wiring and read out equipment requirements prohibited its use from an economic standpoint. This device was a very useful tool in developing stall warning techniques for J47 engines.

Acknowledgments

The author wishes to thank Dr. R. L. Lowry, School of Mechanical Engineering, Oklahoma State University for his contribution to the material presented in this paper. He would also like to thank Mr. W. S. Jones, Deputy Chief, Service Engineering Division, Tinker AFB and his colleagues, specifically, Mr. C. H. Manley and Mr. L. G. Marler whose able assistance and test work made this paper possible.

REFERENCES

1. Kreith, Frank - Principles of Heat Transfer - International Text Book Co. - Scranton, Conn - 1958

2. Wilburn, D. K. - A Survey of Infrared Inspection Measuring Techniques ASTM Materials & Research Standard, Vol I, No. 7 - Jul 1961

3. Beckwith, T. G. and Buck, N. L. - Mechanical Measurements - Addison Wesley Publishing Company, Reading, Mass - 1961

4. Locke, A. S. - Guidance - Van Nostrand Co. - Princeton - 1955

5. Bird, R. B., Steward, W. E. and Lightfood, B. N. - Transport Phenomena - John Wiley & Sons, N. Y. - 1960

DETERMINATION OF STABILITY DERIVATIVES FROM FLIGHT TEST DATA BY MEANS OF OPERATION ANALOG MATCHING

BY

JOHNNY M. RAMPY
1st Lt, USAF

DONALD T. BERRY
Aero Engineer

Air Force Flight Test Center
Edwards Air Force Base, California

(U) DETERMINATION OF STABILITY DERIVATIVES FROM FLIGHT TEST DATA BY MEANS OF HIGH SPEED REPETITIVE OPERATION ANALOG MATCHING

ABSTRACT

The time required and effort required to determine stability derivatives from flight test data containing complex forcing functions is usually very demanding. This paper describes the application of high speed repetitive operation analog matching to the problem. This technique has proven to be five to ten times faster than other methods in general use.

The feasibility of this technique was established using X-15 flight test data and a three degree of freedom lateral-directional simulation of the aircraft.

Derivatives obtained by this technique agree well with those obtained by other methods. In addition to requiring less time, the technique provides an insight into the vehicle dynamics which enables the engineer to check the validity of the flight test data.

I. INTRODUCTION

Knowledge of aerodynamic stability and control derivatives is indispensable to the planning of initial flights of test programs to insure maximum use of flight time and minimize risk to crew and vehicle. The initial flights are made after extensive planning in which simulators are used to identify regions of danger. Boundaries are established that limit the flight envelope to safe areas with regard to normal acceleration, speed, altitude, and controllability. The characteristics used in the simulation are those determined from wind tunnel tests and other ground facilities.

To minimize the risk to the crew and vehicle, it is desirable to expand the flight corridor of a new vehicle in steps and determine the stability derivatives after each flight. The characteristics obtained from ground tests are replaced by those obtained from flight data, and the next flight is planned according to the trends noted in the previous flight. Thus, the simulation of the vehicle becomes progressively more accurate, and the next flight planned more efficiently and safely.

Typical methods used to obtain stability and control derivatives from flight test data are frequency response (Reference 1), transient response (Reference 2), time vector (Reference 3), curve fitting by least squares (Reference 4), and analog matching (Reference 5). Except for analog matching, these techniques require either unforced (absence of continuous control inputs) oscillation data or expensive, tedious calculations.

Analog matching provides a means of determining stability derivatives that is simple and straightforward and can be used conveniently with flight test data in which the vehicle experiences pilot and/or stability augmentation system inputs. However, conventional analog matching can be time consuming.

II. CONVENTIONAL ANALOG MATCHING

The analog computer is programmed with the vehicle equations of motion so as to provide a mathematical model of the aircraft. Theoretical or wind tunnel values of the stability derivatives are used in the equations of motion as a first estimate. All the basic variables that describe the motion of the aircraft, as well as the associated control surface inputs, are recorded from an actual flight test.

The flight test inputs (e.g., aileron and rudder deflections) are reproduced on function generating equipment and introduced into the mathematical model of the aircraft on the analog computer. The computer calculates a response, and this response is recorded by a strip recorder. The computer response is compared with the actual flight test time history. This comparison is facilitated by plotting the aircraft flight test time histories on clear plastic that is used to overlay the computer response time histories.

The difference in the computer and aircraft response is attributed to an error in the estimate of the values of the stability derivatives. The stability derivatives are changed on the computer using a trial and error process and runs are repeated until a match is obtained between computer and aircraft response. This process is time consuming because a considerable number of individual time histories may be necessary before the influence of the various stability derivatives can be appreciated. Hundreds of individual time histories may be necessary to obtain a match and the cataloging of the various influences becomes difficult.

When matching a computer response to a flight test response, it often is not sufficient to adjust only the stability derivatives. The solutions are sensitive to initial conditions. Experience has indicated that small errors in initial conditions read from flight test records noticeably affect the solution. The effect of initial condition errors is to shift the amplitude or rotate a response time history rather than change its general shape. Errors in initial conditions, consequently, can usually be separated from errors in stability derivatives. Therefore, initial conditions are programmed on handset potentiometers, so that they can be adjusted to aid in matching.

Although the technique is relatively simple and straightforward, hours or days may be spent before a satisfactory match is obtained. The quality of the match depends on the experience of the operator and the quality of the flight test data. If the data contains discrepancies, a considerable amount of time may be spent trying to match data that cannot be matched.

III. HIGH SPEED REPETITIVE OPERATION

High speed repetitive operation greatly facilitates studies which investigate the effect of parameter variations on problem solutions. In this mode of operation, the AFFTC computer obtained solutions 100 times faster than when in regular operation. The computer automatically applied initial conditions, introduced forcing functions, and then computed a solution for a predetermined time interval. At the end of the compute time, recycling occurred automatically. The computer stopped the solution, reset, applied the same initial conditions and forcing functions, and repeated the computation. This sequence was repeated at a rate fast enough to make the solution appear as a stationary wave when displayed on an oscilloscope. A permanent record of the solution was obtained by photographing the oscilloscope.

Problem parameters (stability derivatives and initial conditions) were represented by potentiometers that could be adjusted while the computer was operating in the high speed repetitive mode. Thus, a change in a stability derivative would cause a different solution to appear immediately on the oscilloscope.

IV. APPLICATION OF REPETITIVE OPERATION TO THE DETERMINATION OF STABILITY DERIVATIVES

The computer programming and set-up procedures for high speed repetitive operation were the same as for normal computer operation except for a few additional considerations. A convenient length of flight test record had to be chosen, a computation interval selected, and a correspondence between the two established.

A flight test record of 6 to 10 seconds was usually chosen for analysis. Such a record was long enough to display meaningful responses, yet short enough for dynamic pressure and angle-of-attack to be adequately represented by average values, even during re-entry.

In the high speed repetitive operation mode, the computer was capable of compute time intervals from 10 to 80 milliseconds. However, in order to utilize the full width of the display scope for a solution, switch-selected compute-time intervals of 10, 20, 40, or 80 milliseconds had to be used. Other compute-time intervals were obtained by a vernier adjustment that reduced the width of display on the scope.

The following discussion illustrates how the time interval of the flight test record was correlated with the compute-time interval of the high speed repetitive operation mode. For a complete discussion of time scaling analog computers, see References 6 and 7.

The relationship between the compute-time interval and the length of the flight test record can be expressed as

$$\gamma = \beta t \qquad (1)$$

β = time scale factor. This is the ratio of the time required for the computer to calculate a response to the time required for the same response in flight.

t = time length of the flight test record.

γ = time required for the computer to calculate a response equivalent to the flight test record.

β can be considered to be composed of two factors.

$$\beta = \beta_A \beta_B$$

β_A = time scale automatically introduced by high speed repetitive operation mode.

β_B = time scale programmed on the patch board.

In this investigation, $\beta_A = 0.01$. Therefore, equation (1) becomes

$$\gamma = 0.01\beta_B t \qquad (2)$$

To utilize the full width of the display scope, it was necessary to choose β_B and t so the γ would equal a switch-selected compute time. If an 8-second flight test record was used, no additional time scaling on the patch board was necessary ($\beta_B = 1$) as this directly resulted in a γ equal to a switch-selectable value, e.g.:

$$\gamma = (0.01)(1)(8) = 80 \text{ milliseconds}$$

If it desired to utilize a 10-second flight test record, and use 80 milliseconds of compute time, a β_B of 0.8 was necessary.

$$\gamma = (0.01)(0.8)(10) = 80 \text{ milliseconds}$$

In a similar fashion, convenient time scales and compute intervals were chosen for the various flight test records used.

With the time scales established, plastic overlays of the flight data were plotted to a corresponding scale.

The equations of motion used in the study were the linearized lateral-directional equations:

$$\dot{p} = L_\beta \beta + L_p p + L_r r + L_{\delta_a} \delta a + L_{\delta_r} \delta r + \frac{I_{xz}}{I_x} \dot{r}$$

$$\dot{\beta} = Y_\beta \beta + Y_{\delta_a} \delta a + Y_{\delta_r} \delta r + p \sin \alpha$$

$$- r \cos \alpha + \phi \frac{g}{V} \cos \alpha$$

$$\dot{r} = N_\beta \beta + N_p p + N_r r + N_{\delta_a} \delta a + N_{\delta_r} \delta r + \frac{I_{xz}}{I_z} \dot{p}$$

The symbols of the above equations are defined in the Appendix. The equations of motion are derived in References 8 and 9. Lateral-directional equations of motion were used in this study but the technique is suitable for any set of equations.

The flight test data were obtained from the X-15. Wind tunnel estimates of X-15 stability derivatives were used (Reference 10), and actual X-15 flight data were matched. The control inputs, δ_a(aileron) and δ_r(rudder) were programmed on diode function generators. (Diode function generators have a high frequency response which is necessary for high speed operation.)

Initial conditions were obtained from the flight test data and set into the problem.

The computer solution was displayed on a cathode ray tube capable of displaying up to eight variables. In this study, four variables, roll rate, yaw rate, sideslip angle, and roll angle were displayed for matching. The flight test overlay was affixed to the face of the display scope and stability derivative potentiometers adjusted until a match was obtained.

Figure 1 shows a picture of the computer, display scope, overlay and manually operated potentiometers.

A typical example of a matching sequence is shown in Figures 2 through 8, Appendix II. Figure 2 shows a comparison of the computer response using wind tunnel stability derivatives and the actual flight test data. The computer response and flight test response differed markedly in the period of the yaw rate and sideslip motions. The roll rate traces are also dissimilar.

In Figure 3, the value of C_{n_β} (static directional stability derivative) has been decreased 15 percent. The yaw rate and sideslip traces correspond more closely with the aircraft response, but the roll rate traces still do not match.

Figure 4 shows the result of a 15 percent change in C_{l_β} (dihedral effect derivative). The roll rate traces are more similar.

Figures 5 and 6 depict the changes in computer response due to adjustments of C_{l_p} (roll damping derivative), C_{n_r} (yaw damping derivative), $C_{n_{\delta a}}$ (yaw due to aileron derivative) and C_{l_r} (roll due to yaw derivatives).

Figure 7 shows the final match that was obtained. Figure 8 is a comparison of the computer response with the original (wind tunnel) derivatives, aircraft response, and the final match. Figures 9 and 10 are matches obtained with other X-15 flight records.

V. DISCUSSION OF RESULTS

Previous experiences with conventional analog matching indicated that at least 6 hours were usually spent to obtain an acceptable match of computer and flight test responses. However, it was not unusual to spend 16 to 40 hours. During this study, using repetitive operation the maximum time spent on a set of data was 2 hours. Some matches were obtained in 20 minutes. Two hours was felt to be a representative upper limit, since all participants felt that if a satisfactory match wasn't obtained in this time, a match was impossible.

Verification of this fact was evidenced by the following experience. An attempt was made to match three flight records from a particular X-15 flight. After 2 hours work on each set, it was concluded that some undetermined deficiency must exist in the data. An independent agency attempted to match the same data by conventional analog matching methods. They also concluded the data were deficient. However, 2 days were spent on each set of data before this conclusion was reached.

This illustrated another advantage of the repetitive operation technique. By watching changes in response on the scope as stability derivatives were varied, the operator quickly learned the influence of the various derivatives on the response characteristics of the vehicle. This enabled an evaluation of the data that was being analyzed. If all reasonable variations in stability derivatives failed to provide a match, it could be an indication of an error in the data. The analog computer solution was particularly informative when all but one of the response variables could be matched. In such cases, the inconsistency in the flight test data was easily evaluated. In several instances during this study, reversed polarity of instrument calibrations and incorrect fairing of flight data were discovered in this way.

The study verified well known shortcomings of analog matching techniques. Usually only the major derivatives could be determined. Confidence in the results was a function of how well the data were matched and engineering judgment. At times, a good match could not be obtained for unknown reasons. At other times a match may only be fair and results inconclusive. Also, a good match may not be the result of a unique set of stability derivatives. For these reasons, it would not be wise to depend entirely on analog matching for stability derivative determination in a program. Other means should also be employed to verify trends. Despite these shortcomings, analog matching has proved quite useful in practice as, for example, in the X-15 Program (Reference 5).

The use of high speed repetitive operation minimizes the shortcomings of analog matching. It greatly speeds up the process and gives the engineer a better opportunity to evaluate the data and the results.

VI. CONCLUSIONS

The high speed repetitive operation analog matching techniques for determining stability derivatives from flight test data was 5 to 10 times faster than ordinary analog matching. Only lateral-directional derivatives were determined, but longitudinal derivatives could be determined just as quickly and easily. The technique also gave the user a better insight into the vehicle dynamics which was important in evaluating the test data and computer results.

The technique can be used to analyze data in which only control pulses were experienced or data obtained from flights in which the vehicle was constantly controlled. This feature makes it suitable for the analysis of conventional aircraft, high speed aircraft, V/STOL aircraft, and lifting re-entry vehicles. Additionally, it can be used to salvage test data where inadvertent inputs disturbed the planned test maneuvers.

VII. REFERENCES

1. Donegan, James J., Robinson, Samuel, W. Jr., and Gates, Ordway B. Jr., "Determination of Lateral-Stability Derivatives and Transfer-Function Coefficients from Frequency Response Data for Lateral Motions," NASA TN 9383, May 1954.

2. Huff, W. W. Jr., "Application of Dynamic Testing Procedures to Stability and Control Flight Test Programs," NATO Report 191, April 1958.

3. Breuhaus, W. O., "Resume of the Time Vector Method as a Means for Analyzing Aircraft Stability Problems," WADC TR 52-299, November 1962.

4. Shinbrot, Marvin, "A Least Squares Curve Fitting Method with Application to the Calculation of Stability Coefficients from Transient Response Data," NASA TN 2341, April 1951.

5. Yancey, Roxanah B., Rediess, Herman A., and Robinson, Glenn H., "Aerodynamic-Derivative Characteristics of the X-15 Research Airplane as Determined from Flight Tests for Mach Numbers From 0.6 to 3.4," NASA TN D-1060, January 1962.

6. Anon., "Operation of the PACE 231R Computer," Electronics Associates, Inc., Copyright, 1961.

7. Johnson, Clarence S., "Analog Computer Techniques," New York, McGraw-Hill Book Company, Inc, 1956.

8. Berry, Donald T., "Derivation of the Six Degree of Freedom Equations of Motion of an Airplane," AFFTC TIM-62-1056, April 1962.

9. Etkin, Bernard, "Dynamics of Flight, Stability and Control," New York, John Wiley and Sons, Inc., 1959.

10. Anon., "Revised Basic Aerodynamic Characteristics of the X-15 Research Airplane," Rep. NA-59-1203, North American Aviation, Inc., 1959.

APPENDIX I

Nomenclature

Symbol	Definition
b	wing span, ft
c	mean aerodynamic chord, ft
C_l	rolling moment coefficient
C_{l_p}	change in rolling moment coefficient with change in rolling velocity, $\partial C_l/\partial(pb/2V)$, per radian
C_{l_r}	change in rolling moment coefficient with change in yawing velocity, $\partial C_l/\partial(rb/2V)$, per radian
C_{l_β}	change in rolling moment coefficient with change in sideslip angle, $\partial C_l/\partial\beta$, per radian
$C_{l_{\delta a}}$	change in rolling moment coefficient with change in aileron deflection, $\partial C_l/\partial\delta a$, per radian
$C_{l_{\delta r}}$	change in rolling moment coefficient with change in rudder deflection, $\partial C_l/\partial\delta r$, per radian
C_n	yawing moment coefficient, yawing moment/qSb
C_{n_p}	change in yawing moment coefficient with change in rolling velocity, $\partial C_n/\partial(pb/2V)$, per radian
C_{n_r}	change in yawing moment coefficient with change of yawing velocity, $\partial C_n/\partial(rb/2V)$, per radian
C_{n_β}	change in yawing moment coefficient with change in sideslip angle, $\partial C_n/\partial\beta$, per radian

$C_{n_{\delta a}}$ change in yawing moment coefficient with change of aileron deflection, $\partial C_n/\partial \delta a$, per radian

$C_{n_{\delta r}}$ change in yawing moment coefficient with change in rudder deflection, $\partial C_n/\partial \delta r$, per radian

C_y side force coefficient

C_{y_β} change in side force coefficient with change in sideslip angle, $\partial C_y/\partial \beta$, per radian

I_x moment of inertia about longitudinal axis, slugs-ft^2

I_y moment of inertia about lateral axis, slugs-ft^2

I_z moment of inertia about vertical axis, slugs-ft^2

I_{xz} product of inertia, slug-ft^2

L rolling moment

L_p change in rolling moment with change in rolling velocity, $qSb^2C_{l_p}/2VI_x$, per sec

L_r change in rolling moment with change in yawing velocity, $qSb^2C_{l_r}/2VI_x$, per sec

L_β change in rolling moment with change in sideslip angle, $qSbC_{l_\beta}/I_x$, per sec^2

$L_{\delta a}$ change in rolling moment with change in aileron deflection, $qSbC_{l\delta_a}/I_x$, per sec^2

$L_{\delta r}$ change in rolling moment with change in rudder deflection, $qSbC_{l\delta_r}/I_x$, per sec^2

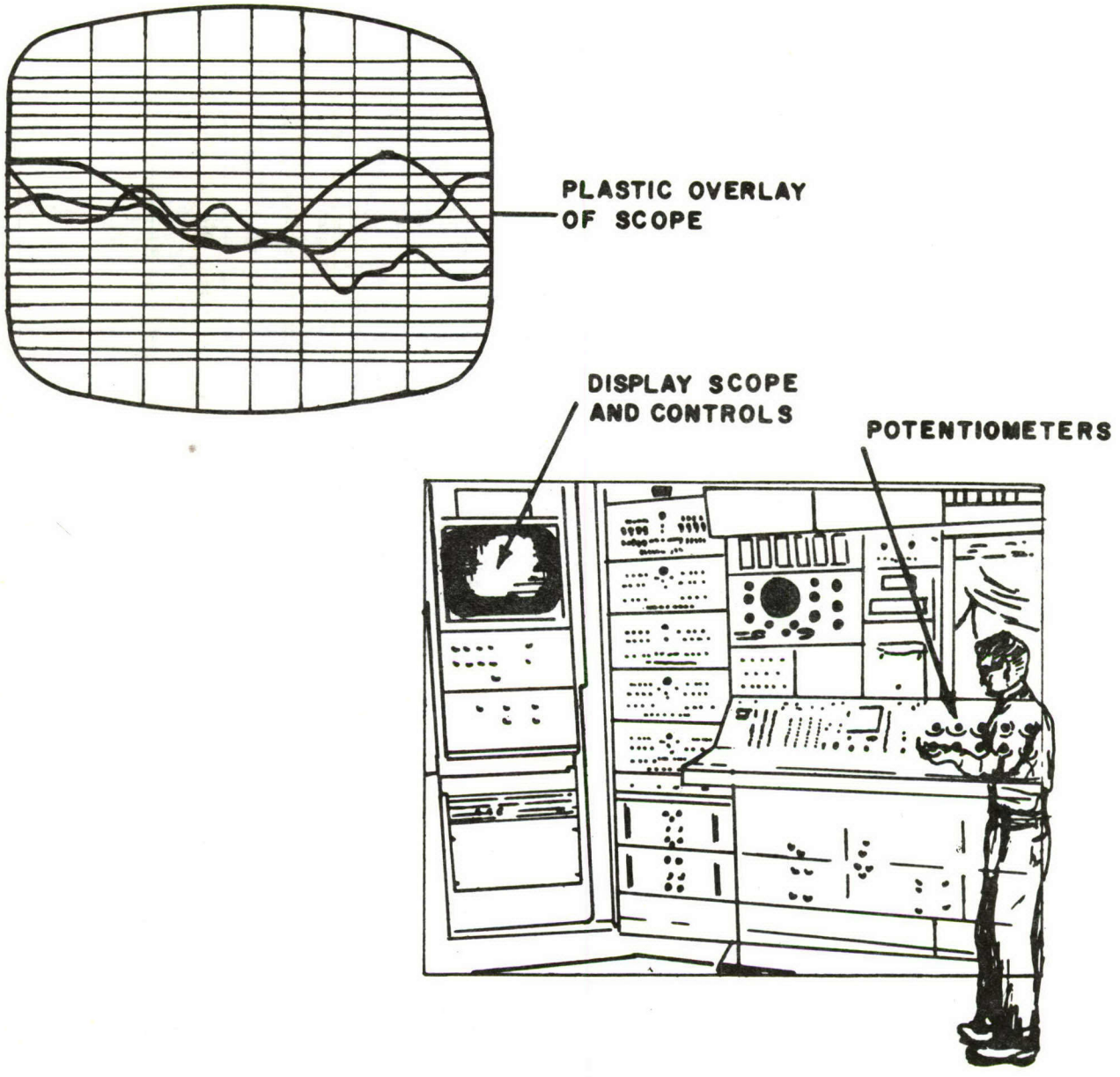
M Mach number

N yawing moment

N_p change in yawing moment with change in rolling velocity, $qSb^2C_{n_p}/2VI_z$, per sec

N_r	change in yawing moment with change of yawing velocity, $qSb^2C_{n_r}/2VI_z$, per sec
N_β	change in yawing moment with change in sideslip, $qSbC_{n_\beta}/I_z$, per sec^2
$N_{\delta a}$	change in yawing moment with change of aileron deflection, $qSbC_{n_{\delta a}}/I_z$, per sec^2
$N_{\delta r}$	change in yawing moment with change in rudder deflection, $qSbC_{n_{\delta r}}/I_z$, per sec^2
p	rate of roll, radians/sec
$\dot{p}$	roll acceleration, radians/sec^2
q	free stream dynamic pressure, lb/sq ft
r	rate of yaw, radians/sec
$\dot{r}$	yaw acceleration, radians/sec^2
S	reference area, sq ft
t	time, sec
V	free stream velocity, ft/sec
α	angle of attack
β	angle of sideslip, radians
$\dot{\beta}$	time rate of change of sideslip angle, radians/sec

FIGURE I
COMPUTER AND REP-OP EQUIPMENT



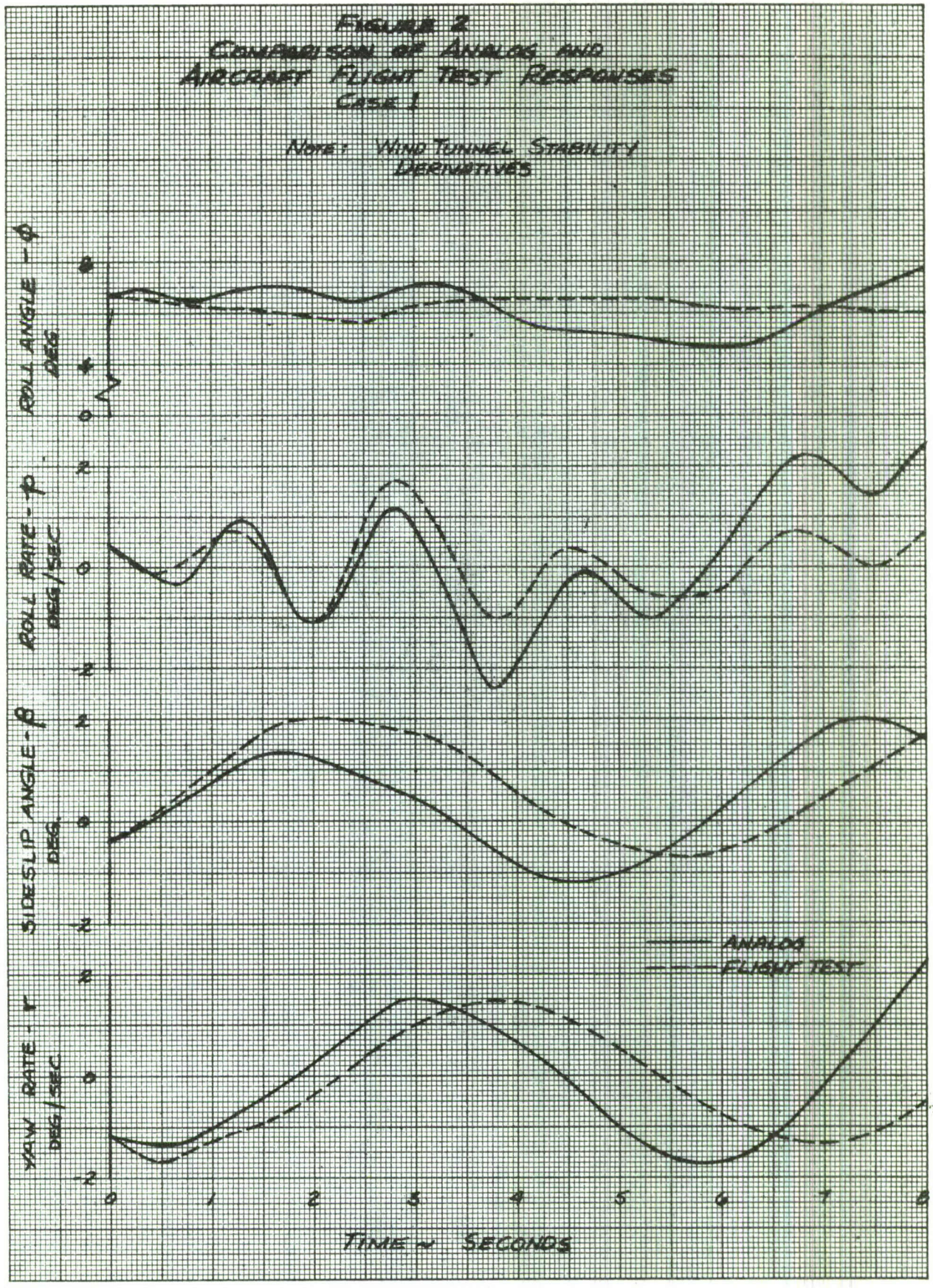

FIGURE 2
COMPARISON OF ANALOG AND AIRCRAFT FLIGHT TEST RESPONSES
CASE 1
NOTE: WIND TUNNEL STABILITY DERIVATIVES
ROLL ANGLE - φ DEG
ROLL RATE - p DEG/SEC
SIDESLIP ANGLE - β DEG.
YAW RATE - r DEG/SEC
ANALOG
FLIGHT TEST
TIME - SECONDS

FIGURE 3
COMPARISON OF ANALOG AND
AIRCRAFT FLIGHT TEST RESPONSES
CASE 1
NOTE: 15% DECREASE IN C_{n_β}

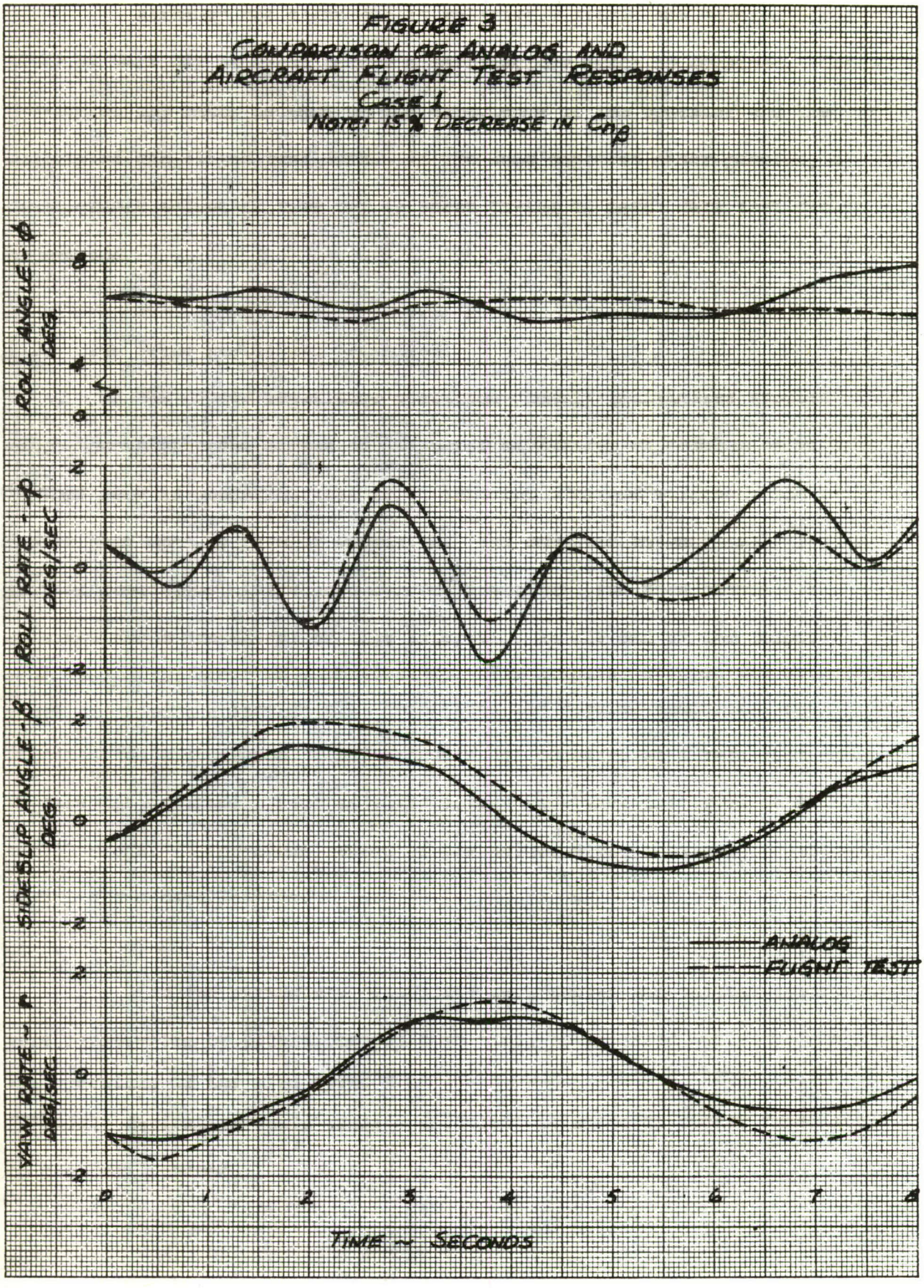
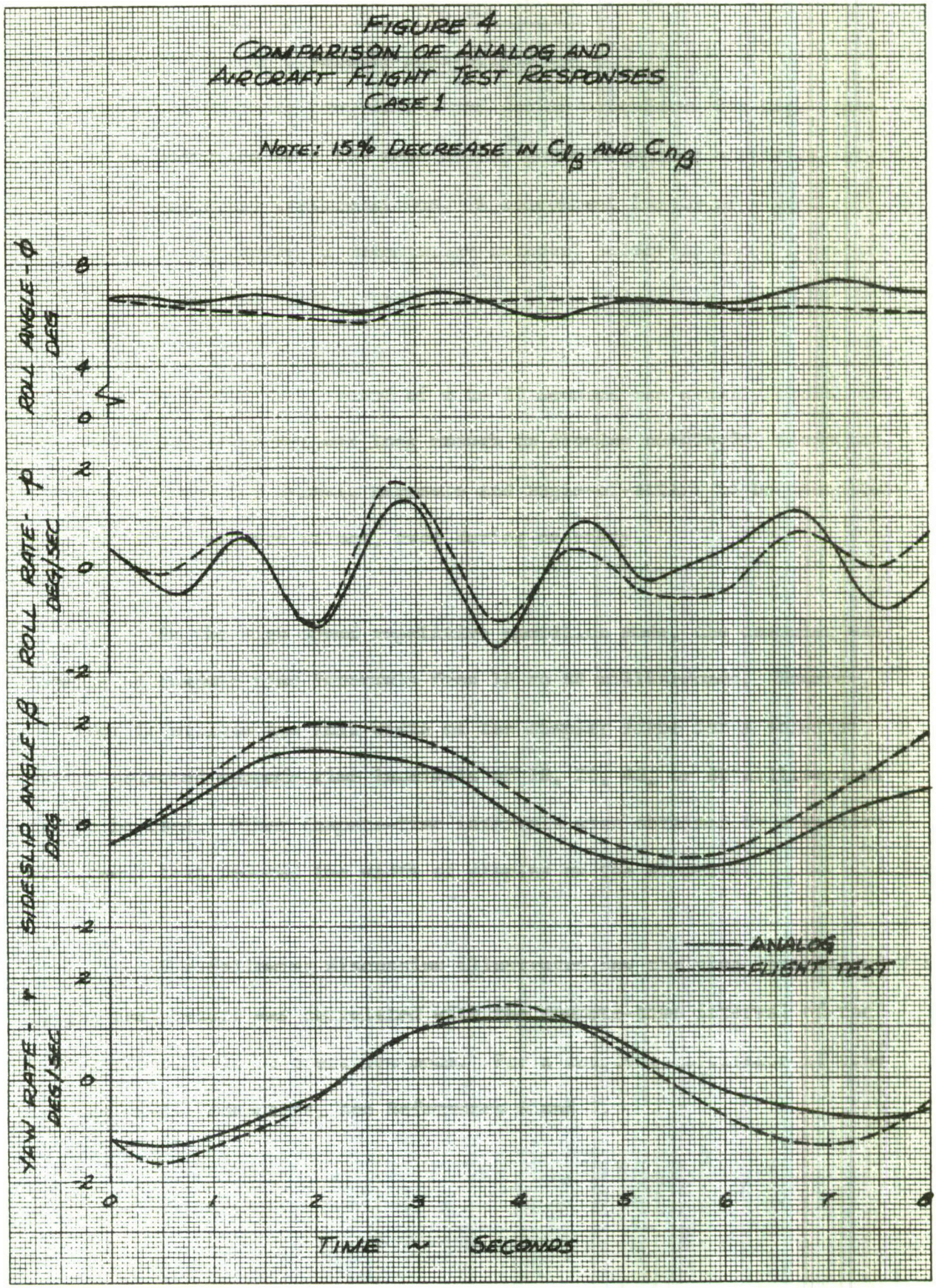
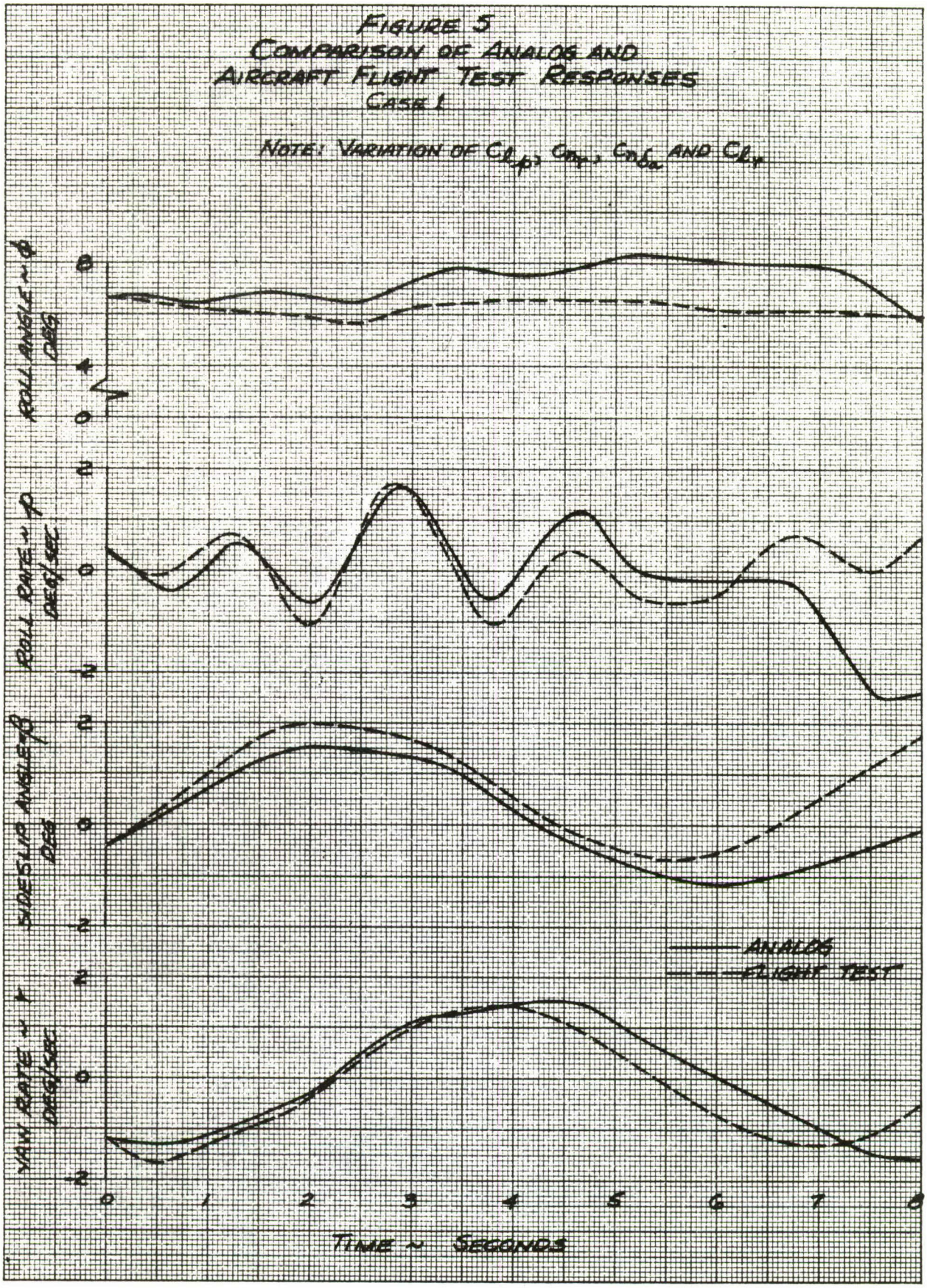
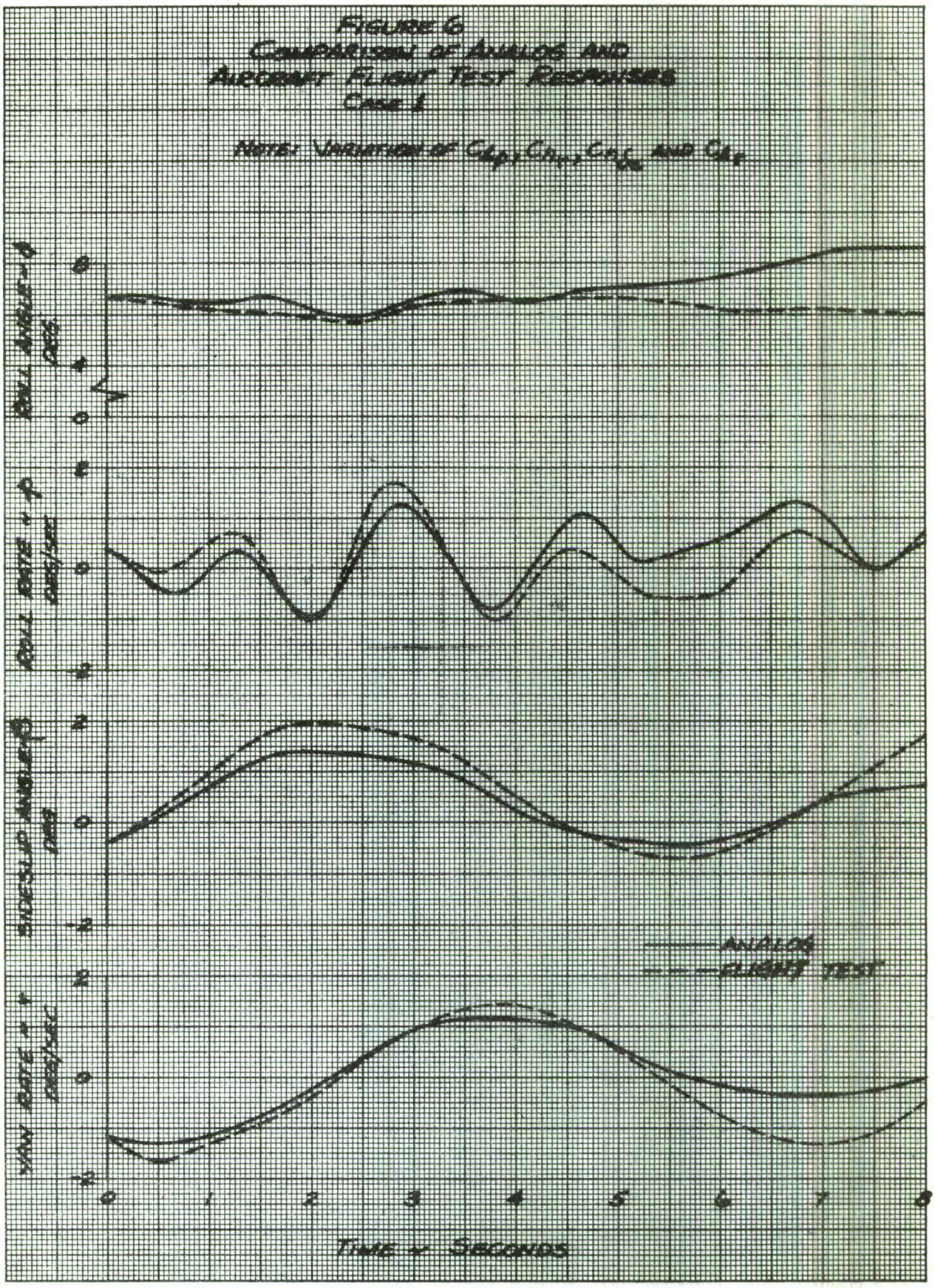


FIGURE 4
COMPARISON OF ANALOG AND
AIRCRAFT FLIGHT TEST RESPONSES
CASE 1

NOTE: 15% DECREASE IN Cl_β AND Cn_β




FIGURE 5
COMPARISON OF ANALOG AND
AIRCRAFT FLIGHT TEST RESPONSES
CASE 1
NOTE: VARIATION OF C_{l_p}, C_{n_r}, $C_{n_{\delta_a}}$ AND C_{l_r}
ROLL ANGLE ~ ϕ DEG
ROLL RATE ~ p DEG/SEC
SIDESLIP ANGLE β DEG
YAW RATE ~ r DEG/SEC
ANALOG
FLIGHT TEST
TIME ~ SECONDS


FIGURE 6
COMPARISON OF ANALOG AND AIRCRAFT FLIGHT TEST RESPONSES
CASE 1
NOTE: VARIATION OF C_{l_p}, C_{n_r}, $C_{n_{\delta_a}}$ AND C_{l_r}
ROLL ANGLE, φ, DEGREES
ROLL RATE, p, DEGREES/SEC
YAW RATE, r, DEGREES/SEC
SIDESLIP ANGLE, β, DEGREES
ANALOG
FLIGHT TEST
TIME - SECONDS
0
1
2
3
4
5
6
7
8

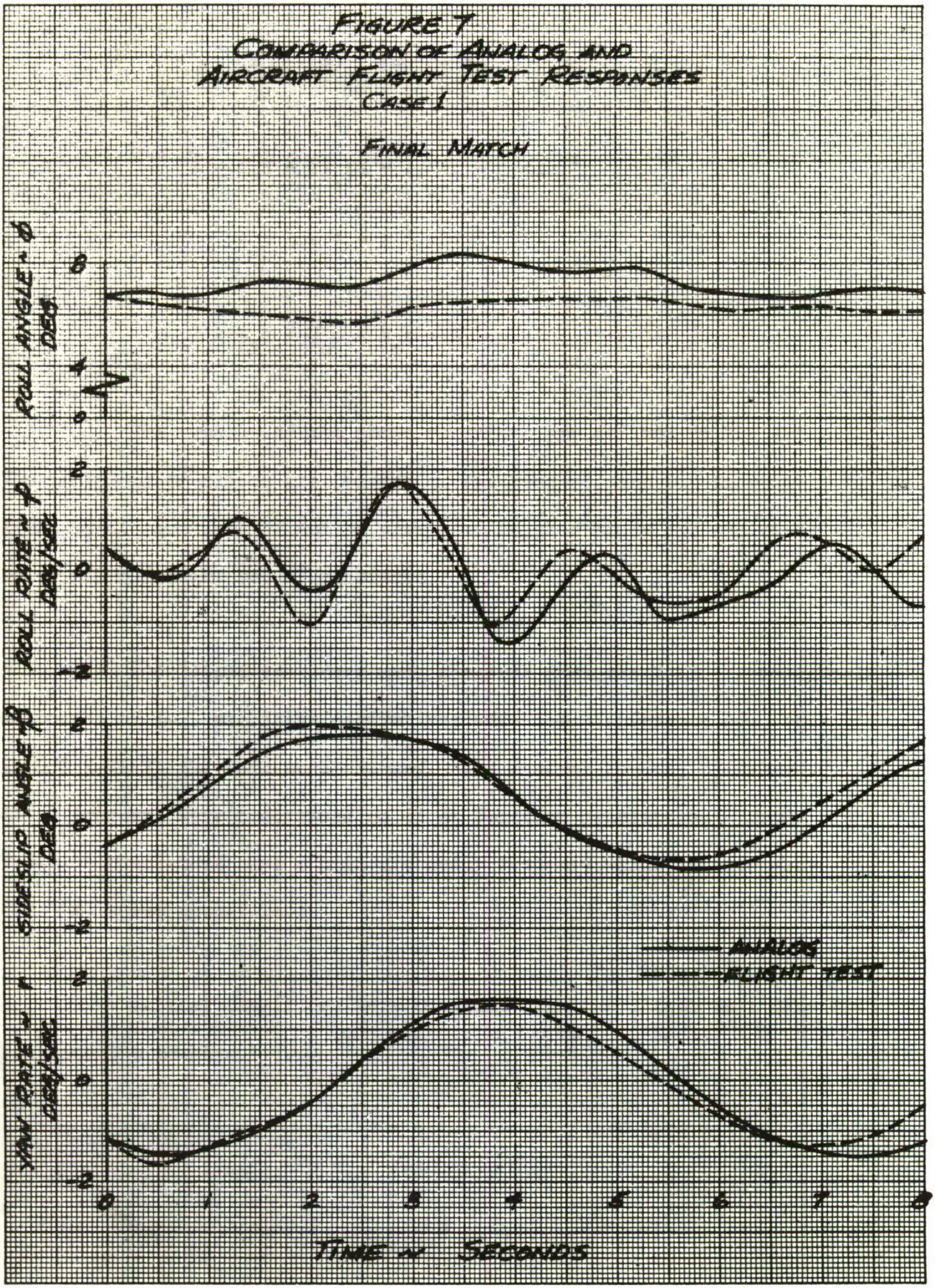
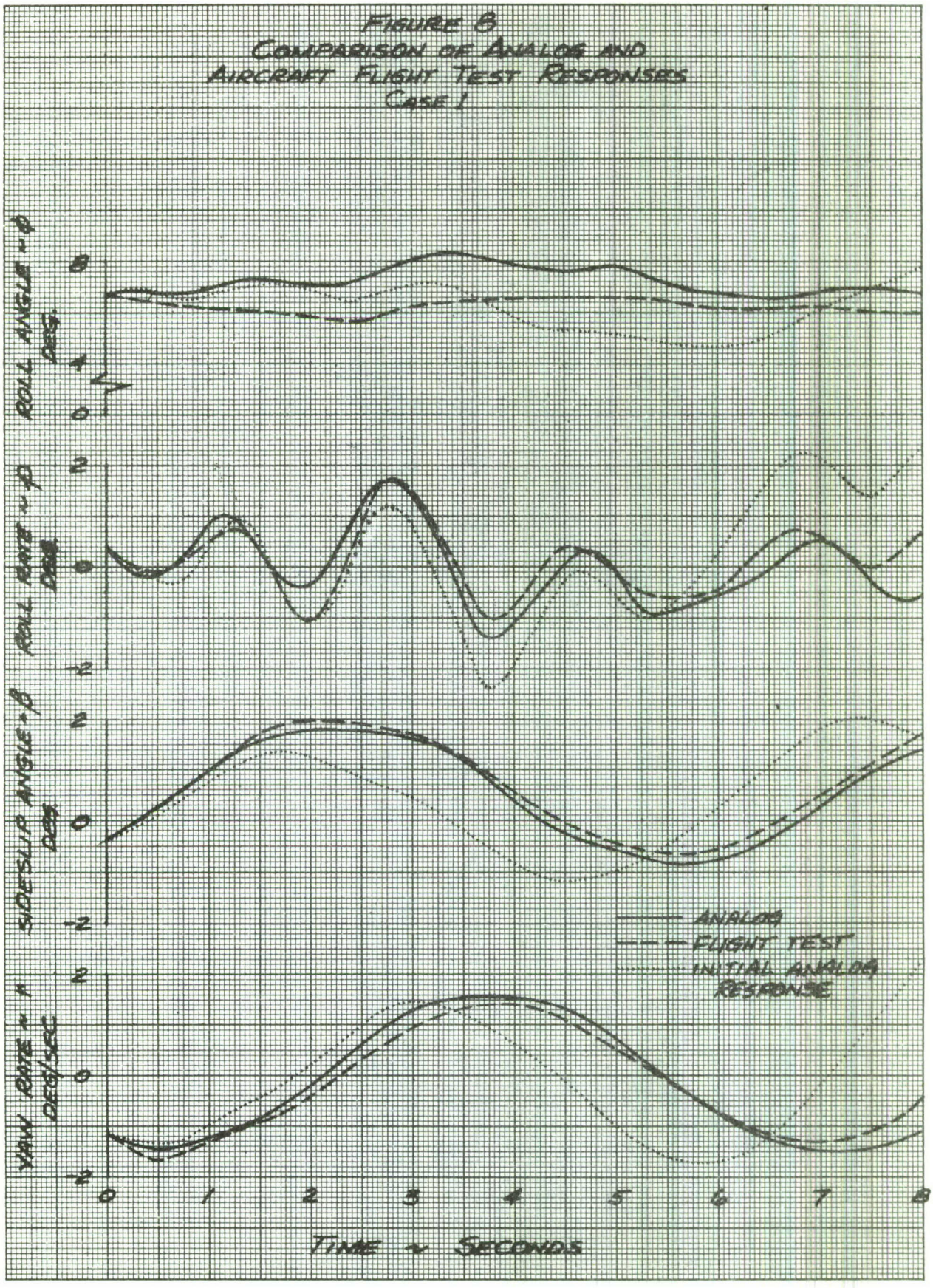
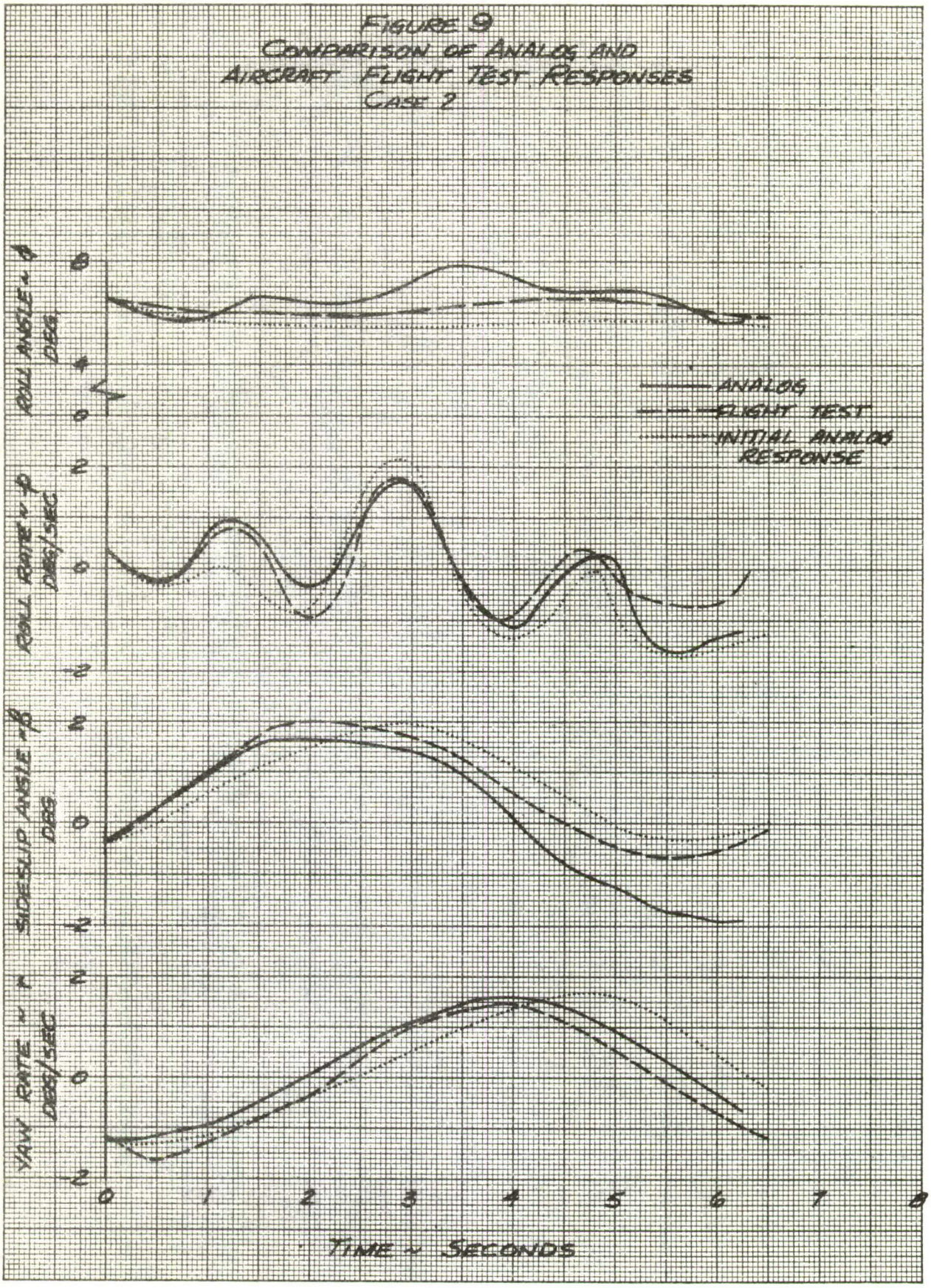
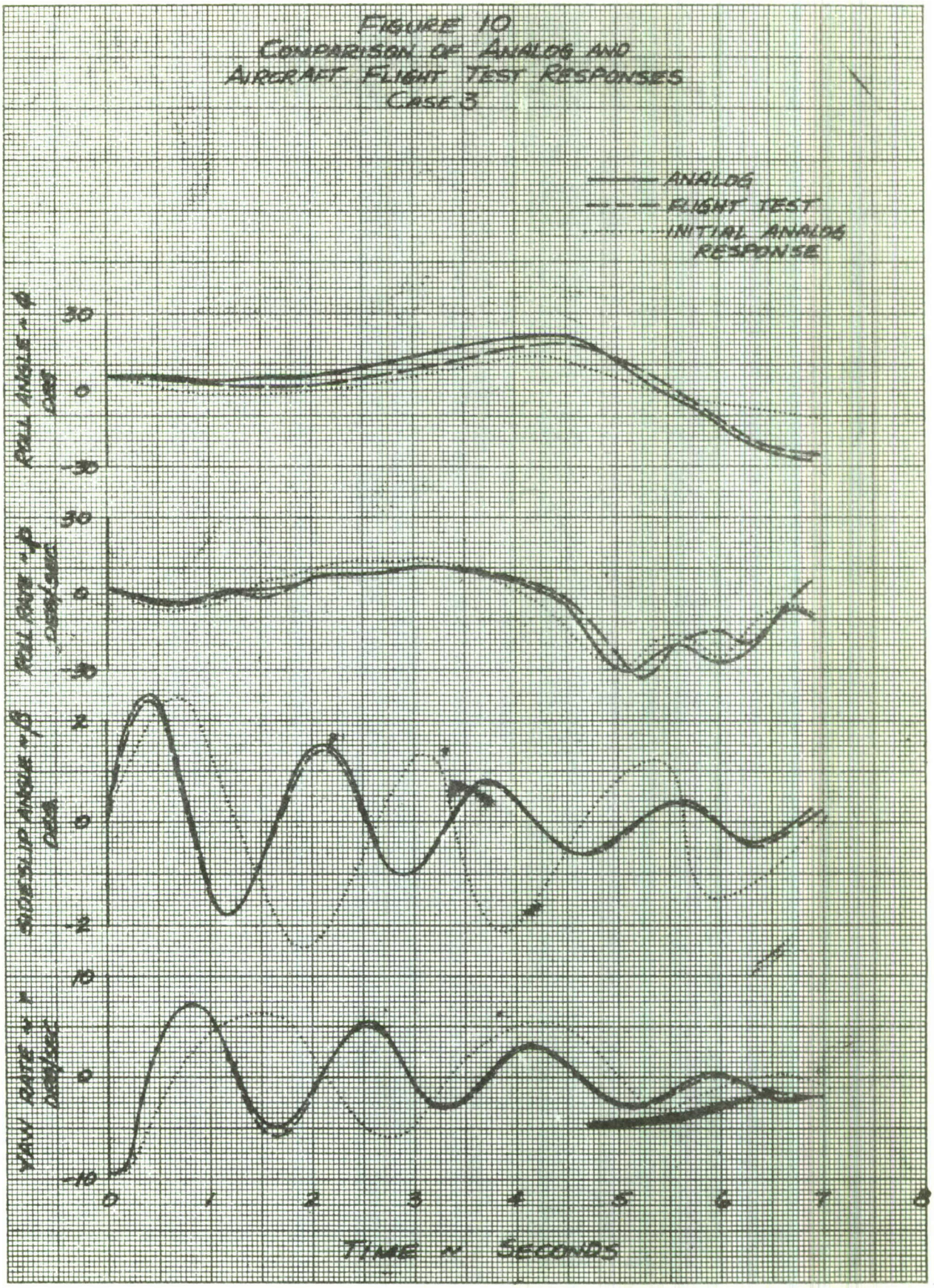

FIGURE 7
COMPARISON OF ANALOG AND
AIRCRAFT FLIGHT TEST RESPONSES
CASE 1
FINAL MATCH
ROLL ANGLE, φ DEG
8
4
0
-4
ROLL RATE, p DEG/SEC
2
0
-2
SIDESLIP ANGLE, β DEG
4
0
-4
YAW RATE, r DEG/SEC
2
0
-2
ANALOG
FLIGHT TEST
0
1
2
3
4
5
6
7
8
TIME IN SECONDS

FIGURE 8
COMPARISON OF ANALOG AND
AIRCRAFT FLIGHT TEST RESPONSES
CASE 1




FIGURE 9
COMPARISON OF ANALOG AND
AIRCRAFT FLIGHT TEST RESPONSES
CASE 2
ROLL ANGLE - Φ
DEG
ROLL RATE - P
DEG/SEC
SIDESLIP ANGLE - β
DEG
YAW RATE - r
DEG/SEC
ANALOG
FLIGHT TEST
INITIAL ANALOG
RESPONSE
TIME - SECONDS


FIGURE 10
COMPARISON OF ANALOG AND
AIRCRAFT FLIGHT TEST RESPONSES
CASE 3
ANALOG
FLIGHT TEST
INITIAL ANALOG
RESPONSE
ROLL ANGLE, Φ
30
0
-30
PITCH ANGLE, θ
30
0
-30
YAW RATE, r
2
0
-2
YAW ANGLE, ψ
10
0
-10
0
1
2
3
4
5
6
7
8
TIME IN SECONDS

(U) RADIOISOTOPE TECHNIQUES FOR THE MEASUREMENT OF VTOL LANDING AND TAKEOFF PARAMETERS

by

Paul Polishuk
and
1st Lt. Rudy C. Beavin

AIR FORCE FLIGHT DYNAMICS LABORATORY
RESEARCH & TECHNOLOGY DIVISION
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

ABSTRACT

The problem of control data acquisition during takeoff and landing of VTOL type vehicles is explored. From a brief survey of the required data inputs and performance data for a tactical VTOL aircraft, a tentative landing profile has been established. Of the possible phenomena, such as optical, microwave, and nucleonic which may be used for measurement of these landing parameters, only several nucleonic techniques are analyzed in this paper. The radioisotope systems described and analyzed herein all consist of ground sources and detector packages located in the vehicle. They include rotating collimator, servoed, matrix, intensity, energy discrimination, and pulse modulation systems. The advantages of the radioisotope systems are that they involve passive ground sources, (i.e., no ground power requirement and no radio, thermal, or optical emission), have all weather capability, can be made mobile, and have low susceptibility to jamming. Of the systems analyzed, the servoed and rotating collimator systems look the most promising; however, further experimental work is needed in determining build-up factors, accuracy and speed of response of angle measurements, effects of random motion of servoed systems, and angle measurement by rotating collimators. Although the systems analyzed do not presently give a clear cut solution to the data acquisition problem during the terminal phase of VTOL landing, subsequent experimental work may make such a solution possible. It must also be recognized that radioisotope systems other than those presented in this report may be conceived which do not have the attendant problems discussed above.

TABLE OF CONTENTS

Para.	Title	Page
1.0	INTRODUCTION	1
1.1	NUCLEONIC TECHNIQUES	4
1.2	DETECTION OF GAMMA RAYS	6
1.3	ATTENUATION OF GAMMA RAYS	7
1.4	RADIATION HAZARD	8
2.0	LANDING PROFILE AND SYSTEMS DEFINITIONS	9
2.1	DISCUSSION OF A POSSIBLE LANDING PROFILE FOR A TYPICAL VTOL VEHICLE AND SELECTION OF LIMITS OF RADIOISOTOPE SYSTEM LANDING PROFILE	9
2.1.1	POSSIBLE LANDING PROFILE	9
2.1.2	LIMITS OF RADIOISOTOPE SYSTEMS STUDIED	10
2.2	SYSTEMS DEFINITIONS	11
3.0	DISCUSSION OF TECHNIQUES	14
3.1	ROTATING COLLIMATOR SYSTEMS	14
3.2	SERVOED SYSTEMS	18
3.3	MATRIX SYSTEMS	22
3.4	INTENSITY MEASURING SYSTEMS	23
3.5	ENERGY DISCRIMINATION	25
3.6	INTENSITY MODULATION	25
4.0	CONCLUSIONS	26
5.0	RECOMMENDATIONS	28

LIST OF ILLUSTRATIONS

Fig. Nr.	Title	Page
1.	TYPICAL LANDING PROFILE FOR TACTICAL VTOL	32
2.	REQUIRED SOURCE STRENGTHS AS A FUNCTION OF RANGE	33
3.	FORWARD VTOL SITE	34
4.	GROUND PLANE AND ELEVATION ANGLE MEASUREMENT SYSTEM GEOMETRY	35
5.	GROUND PLANE AND LINE OF SOURCES PLANE ANGLE MEASUREMENT SYSTEM GEOMETRY	36
6.	REFERENCE PLANE ANGLE MEASUREMENT SYSTEM GEOMETRY	37
7.	EGG CARTON COLLIMATOR	38
8.	BLINDED DETECTOR-COLLIMATOR PACKAGE	39
9.	SERVOED SYSTEM ANGLE MEASUREMENT GEOMETRY	40
10.	ABSOLUTE INTENSITY MEASUREMENT SYSTEM GEOMETRY	41
11.	ROTATING COLLIMATOR-DETECTOR PACKAGE	42
12.	MATRIX PATTERN FOR CONVENTIONAL AIRCRAFT	43

TABLES

Table Nr.	Title	Page
I	COMPARISON OF USABLE PHENOMENA FOR VTOL LANDING	3
II	PRESENT AND FUTURE SPECIFIC ACTIVITIES OF Co-60	6

1.0 INTRODUCTION

There exists a critical problem associated with vertical takeoff and landing aircraft, (VTOL), which is the measurement of position and position rates during the terminal phase of flight. Since the Air Force background in the design of, flight test of, and operational experience with these types of vehicles is rather limited, there exists no known proven technique for the measurement of these parameters. In addition, it may be necessary to measure relative vehicle orientation with respect to a set of reference coordinates, range from a prescribed landing area, glide slope angle, etc. New methods for the measurement of these parameters must be developed to provide the necessary information during this very critical phase of flight, i.e., terminal phase. A typical landing profile for a tactical VTOL is shown in Fig. 1 (Ref I). As can be seen, a transition from horizontal to vertical flight is required which is a departure from conventional landings and presents a host of new measurement problems.

The measurement problems for the terminal phase of flight are considered to be provision of accurate position and position rate information, referenced to a coordinate system situated on the ground plane. It is recognized that an attendant problem is that of bringing the aircraft within the range of any proposed terminal phase system. A problem unique to VTOL aircraft is that the engines will be operated in a vertical mode. This may cause a great deal of dirt, sand, and other debris to be thrown into the air by the engine exhaust interacting with the ground. Any

proposed landing system must be capable of operating independently of these effects. Desirable features of a system for measuring ground reference data for flight control use during the approach, transition, hover, and descent to touchdown portions of the tactical VTOL aircraft mission include:

a. All weather capability.

b. Passive ground components.

c. Low susceptibility to jamming.

d. Simplicity and reliability in operation, maintenance and construction.

e. Light weight, transportability.

f. Low power consumption.

g. Presentation of minimum hazard to operating and using personnel.

h. Long life and maintenance free operation.

i. Provision of useful information for approach, hover, and touchdown phases of terminal flight.

j. Freedom from difficult or complex alignment problems.

There are several phenomena which might be used for the measurement of the required landing parameters and which embody most or all of the features stated above. These include optical and infrared, radar and microwave, and X-ray and nucleonic techniques. Each of these has inherent problems as shown in Table I (Ref 2). Since all the phenomena described in Table I are electromagnetic and nucleonic in nature, their attenuation in air and other materials obeys the exponential law, ($I \propto e^{-R/\lambda}$ where, I is intensity, R is path distance and λ is the characteristic length). The characteristic length as mentioned in Table I is defined

TABLE I - COMPARISON OF USABLE PHENOMENA
FOR VTOL LANDING MEASUREMENT

Principal Phenomena	Characteristic Length in Air (Foul Weather)	Possible Max. Transmission	Detector	Source and Associated Equipment	Remarks
Optical (Visual)	Less than 50' Fog	Miles	Well developed art. Small lightweight & efficient (5 to 10%)	Random and Coherent Large sources Tends toward heavy.	Will not penetrate fog. Not secure. Detector on every enemy soldier.
Infrared	Less than 100' Fog	Miles	Well developed art. Lightweight	Random and Coherent Tends toward heavy.	With high power laser might penetrate but would require complex equipment.
Light Ultra-Violet	Less than 30' Fog	Miles	Not as well developed as other light frequencies.	Random. Tends toward heavy.	Will not penetrate fog.
Microwave	Limited in heavy rain.	Dozens of miles.	Well developed art. Complex. Medium weight	Coherent Heavy	Complex ground installation or aircraft installation. Not secure. Problem of ground reflection at low altitude.
Gamma Rays	350 - 450 Feet	A few thousand feet.	Well developed. Medium size and weight.	Random shielding. May require heavy material.	Simple ground installation. Simple air installation. Secure. No problem of ground reflection.
X-Rays	200 - 400 feet depending on energy	A few thousand feet	Well developed. Medium size and weight	Random. Large Heavy	Complex ground installation Simple air installation. Secure.
Alpha and Beta Particles	Inches	Inches	Easily detected	Random Small, light-weight.	Not enough range.
Neutrons	Hundreds of feet	A few thousand feet	Detectors are difficult & low in efficiency.	Random Complex	Source and detector arrangement too complex.

as that path distance in air which reduces the radiation intensity to 1/e or 36.8% of its initial intensity. This varies only slightly under all conditions of smoke, fog, and rain. The worst short term perturbation of density occurs during a rainstorm when the air density is increased by a nominal 1/2% (Ref 2). In this case, for gamma rays of about 1 Mev, the characteristic length is modified by only 1/2% (Ref 2). Although all the phenomena listed require careful scrutiny for possible application, only the nucleonic techniques will be discussed here.

The Instrumentation Physics Facility of the Control Elements Branch, Flight Control Division, AF Flight Dynamics Laboratory has been engaged in the investigation of nucleonic techniques for control data measurements for sometime (Ref 3). Experimental facilities are presently available for the evaluation and development of nucleonic techniques for Control Data Measurements in addition to landing such as mass fuel quantity, fuel mass flow, density, position, and acceleration. Based on previous contracted effort (Ref 4), and a basic investigation of the Short Take-Off Landing (STOL) requirements, a proposed design for a radioisotope landing data system for STOL has been proposed (Ref 5). The analytical investigations of this report draw upon this previous experience. Preliminary results of the investigation reveal that there are some problems associated with the radioisotope techniques and these are discussed below. Experimental investigations have been initiated in an attempt to find solutions for these problems.

1.1 NUCLEONIC TECHNIQUES

The following is a brief discussion of the basic factors involved in the application of nucleonic techniques to the measurement of VTOL landing parameters.

In the decay of radioactive nuclei the de-excitation of an excited nucleus can happen in either of several ways with the resultant emission of alpha particles, beta particles, gamma rays or neutrons. Because of their charge, alpha and beta particles have relatively short ranges in air, and hence can be excluded from further discussion as having application to VTOL measurement problems. Gamma rays, on the other hand, being electromagnetic in nature and having no charge, have a relatively long range in air (characteristic length of about 450 feet in air) and hence have potential applications. X-rays and Bremstrahlung are also electromagnetic in nature, however, because of the short characteristic length (210 feet for 200 kev photons), these may be excluded from further consideration. The neutron has a mass about 2000 times that of the electron, but has zero charge and hence a long range in air. Because the production of intense neutron beams is possible only by charged particle accelerators or reactors and neutrons cannot be simply detected, they shall be excluded from further discussion.

Since only gamma rays have application to the problem at hand, it is well to review some of the properties of gamma rays resulting from decay of radioactive nuclei. Two of the most commonly used gamma emitting radioisotopes are Co-60 and Cs-137. Their common usage can be attributed to their being readily available and their having relatively long half lives, 5.36 years and 26.6 years respectively. Rather high intensities can be obtained using small quantities of these

materials. For example, the present and future anticipated specific activities for Cobalt -60 are shown in Table II (Ref 6):

TABLE II

PRESENT AND FUTURE SPECIFIC ACTIVITIES OF Co-60

Present - 200 Curies/gm

1966 - 400 - 500 Curies/gm

Post 1966 - 500 - 1000 Curies/gm

Because of the high energy nature of gamma rays and their correspondingly high penetration power, dense material such as iron, mercury or lead is required to attenuate them. Though it is necessary to shield Co-60, or other gamma emitting nuclei, with material such as lead for personnel protection, it is also possible to generate well defined beams of radiation or to perform directional discrimination utilizing ground based or airborne collimators, respectively. The sharpness of the beam depends only on the source size and collimator dimensions (Refs 4 and 5). In many of the cases enumerated below, source collimation and/or detector collimation plays an important part.

1.2 DETECTION OF GAMMA RAYS

Detection of gamma rays may be performed by any one of the following methods: Geiger-Mueller (GM) tubes, sodium iodide crystals, (NaI(Tl)), and solid state detectors. Because of the low detection efficiency for gamma rays, and the slow recovery characteristics which limit the count rate, GM tubes can be used only where the source strength is relatively large. Solid state detectors such as Lithium drifted silicon can be operated at very high count rates, because of the rapid charge

collection time, however, because of their low detection efficiency ($<1\%$ for 1 Mev gamma rays) solid state detectors require high intensity sources. Sodium iodide crystals on the other hand offer the best potential because of their high detection efficiency for gamma rays ($>30\%$ for 3" x 3" crystals and 1 Mev gamma rays) and their ability to handle high count rates. This superior efficiency makes NaI(Tl) the most logical choice as a detector in the nucleonic VTOL landing applications to be discussed.

1.3 ATTENUATION OF GAMMA RAYS

The attenuation of gamma rays in materials, air in this case, is a function of several variables, namely, the amount of material to be traversed and the linear absorption coefficient of the material. The number of gamma rays sensed by a detector at a distance R from a point source is given by:

$$N = \frac{c\, d\, n\, A\, E\, e^{-\sigma_a R}\, B(\sigma_a R)}{4\pi R^2} \qquad (1)$$

where

N = counts/sec recorded by the detector

C = number of curies

d = disintegrations/sec-curie = 3.7×10^{10}

n = number of gammas/disintegration = 2 for Co-60

A = detector area (ft^2) = 0.0625 for 3" x 3" NaI(Tl)

E = detector efficiency = 0.6 for NaI(Tl)

σ_a = Linear absorption coefficient for air (ft^{-1}) = 2.21×10^{-3}

R = distance from source in feet

B (σaR) = build-up factor

The build-up factor B(σaR) is due to gamma rays from the source being scattered by the air back to the detector. The build-up factor may be expressed empirically as

$$B(\sigma aR) = 1 + a\,\sigma aR + b(\sigma aR)^2 + - - - - \qquad (2)$$

where "a", "b", etc., are constants (Ref 7). When close to the source B(σaR) approaches unity, but in the neighborhood of 500 feet and greater the second term predominates. The value of "a" generally falls between 4 and 7, depending on the primary gamma ray energy and the discriminator setting. For Co-60 gamma rays, the values of "a", using discriminator settings of 50 kev and 100 kev are 5.4 and 4.2 respectively. Figure 2 gives the relative source strength for the detection of 5000 cps and for values of "a", equal to 1 and 5 in equation (1).

1.4 RADIATION HAZARD

Assuming the use of Co-60 in any practical system the dose obtained from a point source is given by:

$$D = \frac{14.1\,A\,e^{-\sigma_a R}\,B'(\sigma_a R)}{R^2} \qquad (3)$$

where

D = dose in roentgens/hour

A = activity of Co-60 in curies

R = distance from the source in feet

B'(σaR) = build-up factor for absorbed Dose $\approx$1 (Ref 7)

As an example of the safety of radioisotope systems, consider a location 100' from 3 twenty-five (25) curie sources (as might be used in one of

the landing systems described below). It is found that the dose rate is only 87mr/hr. Assuming that a landing takes a maximum of 10 minutes, the dose is less than 15mr. Using the maximum permissable exposure per day as specified by the AEC, 150mr/day, a pilot can make ten landings per day and not exceed his allowed dosage. Of major importance in the design of any radioisotope system is the problem of keeping the source strength to a level which will allow complete safety for all concerned. Using proper shielding and precautions it is possible to use radioisotopes for landing with complete safety (Ref 5).

2.0 LANDING PROFILE AND SYSTEM DEFINITIONS

2.1 DISCUSSION OF A POSSIBLE LANDING PROFILE FOR A TYPICAL VTOL VEHICLE AND SELECTION OF LIMITS OF RADIOISOTOPE SYSTEM LANDING PROFILE

2.1.1 POSSIBLE LANDING PROFILE

Since Air Force experience with VTOL type vehicles is very limited, it is not possible to well define a landing profile at this time. However, based on discussions at ASD (Ref 1), one possible landing profile for a tactical VTOL has been determined (Fig 1). It is proposed that a localizer beacon (possibly non-nucleonic) will be provided about 2 miles distant from the landing area from which it will be possible to get heading information to the landing site. Once a heading is established, a gentle descent is initiated to about 2000' from the site where transition from horizontal to vertical flight begins. Directly over the point of touchdown, the altitude is approximately 50' and direct descent to touchdown is initiated.

The landing area is shown in Figure 3. This is proposed as a dispersed site with very little preparation except for terrain leveling and temporary surface conditioning. The landing area is a 200' x 200' region. It is possible that a corridor, free of obstacles greater than 2 feet in height, may be provided to a distance of 1/4 mile. The actual landing site is a circle of 40' diameter where the soil has been treated with some stabilizing agent such as a concrete/resin and/or covered with a membrane. Outside this circle and within the 200' wide area, the ground will be cleared of obstacles of greater than 4" in height. It is postulated that surrounding the landing area there may be trees of heights ranging from 25 feet to 125 feet.

During landing, it is assumed that descent rates may vary from 5 - 16 ft/sec, with lateral motion of about 1 - 5 ft/sec. Altitude and lateral position accuracies may be required within $\pm$ 1 foot.

2.1.2 LIMITS OF RADIOISOTOPE SYSTEMS STUDIED

Those radioisotope landing systems developed and studied by the authors for the VTOL application are discussed under rotating collimator systems. The requirements for accurate position and position rate information is most critical during the very last part of the landing maneuver. For purposes of analysis this last portion has been taken to be from a range of 600 feet and an altitude of 150 feet to touchdown. It is assumed that the localizer beacon will be able to point the aircraft toward the landing area, and bring it to the region under consideration. Extension of the radioisotope systems to further distances from touchdown are possible, however, greater source

strengths would be required since the aircraft would be at a greater range and altitude.

2.2 SYSTEMS DEFINITIONS

Basic to the measurement of position and position rate is the determination of vehicle position with respect to a fixed reference. Position rate, is obtained as a time derivative in the case of continuously measuring systems or as an average velocity over the measurement period in the case of discretely measuring systems, i.e., systems which require a finite time to make a measurement, each successive measurement being independent of that preceding it. The measurement or calculation of the orientation, with respect to a fixed reference system, of the distance vector extending from the vehicle to the desired landing site is seen to be both necessary and sufficient for determination of position and hence position rate; provided the information rate is rapid enough to permit meaningful calculations. The following are brief descriptions of conceptually feasible radioisotopic techniques for determining the radius vector:

a. There are several techniques conceived by the authors which are generically called Rotating Collimator Systems. They are defined below:

(1) Ground Plane and Elevation Angle Measurement - This technique depends on the measurement of the three angles shown as α , β , γ in Figure 4. Knowledge of these angles permits the calculation of the X-Y-Z coordinates of point P and hence the determination of the vector PO.

(2) Ground Plane and Line of Sources Plane Angle Measurements - This technique depends on the measurement of the three angles shown as α, β, γ in Figure 5. Knowledge of these angles permits the calculation of the X-Y-Z coordinates of point P and hence the determination of vector PO.

(3) Reference Plane Angle Measurements - This technique depends on the measurement of the three angles shown as α, β, γ in Figure 6. Knowledge of these angles permits the calculation of the X-Y-Z coordinates of point P and hence the determination of the vector PO.

b. There are two techniques, differing only in collimator design, which may generically be called Servoed Systems. These are the Egg Carton System Collimator conceived by the authors and the Blinded Detector System proposed by Industrial Nucleonics (Ref 2). The airborne collimator-detector packages associated with these two systems are shown in Figure 7 and Figure 8, respectively. Both techniques depend on the measurement of the three angles shown as α, β, γ in Figure 9. It is to be noted that in this case γ is not measured in the ground plane. Knowledge of these angles permits one as before to determine the vector PO.

c. Another group of systems falls into a class which might be termed Matrix Systems. All of these techniques depend on the mapping of the space near the vehicle destination with beams of radiation. It is necessary to depend on being able to detect the persence or absence of radiation and a bookkeeping system to determine position as either within or without a finite volume of space.

d. The measurement of radiation intensity is another possible means of determining position. The two techniques described below fall into this category:

(1) Absolute Intensity - This technique depends on the measurement of the absolute intensity of radiation, i.e., determination of r_1, r_2, and r_3 as shown in Figure 10. In either case the X-Y-Z coordinates of P are determined.

(2) Relative Intensity - This technique must be augmented by other techniques to form a complete system. Here only the relative count rates from two or more sources are determined and hence the ratio of the lengths of the vectors from the detector to each source. This information, coupled with an angular measurement, permits position determination. Without the angular measurement only a "localizer" function is provided, i.e., the vehicle may be directed so as to be equidistant from two sources.

e. There are two more techniques available which may augment those previously discussed. Neither of these will permit the unique determination of position by itself, but in some instances may simplify the logic associated with other systems if incorporated. They are as follows:

(1) Energy Discrimination - Use of this technique permits identification of the source being interrogated. This in general simplifies the logical manipulation of data required from measurement to vector determination. The method depends on recognition of the unique photon energies associated with different radioisotopes.

(2) Intensity Modulation - This technique also permits identification of the source being interrogated. The method this time, however, depends on recognition of variation in intensity and the rate of intensity modulation. Again simplifications in the logic required may be realized.

3.0 DISCUSSION OF TECHNIQUES

3.1 ROTATING COLLIMATOR SYSTEMS

a. Concept - These systems appear promising, since through their use it is possible to measure angles in the X-Y, Y-Z, and Z-X planes formed by a reference system on the ground (Figure 6). This capability greatly reduces the complexity of the measurements. The airborne systems must, however, be mounted on a stable platform aligned with true north and local vertical as determined by the vehicle inertial system.

The following is a discussion of technique 2.2(a-1), Ground Plane and Elevation Angle Measurement (Figure 4). It is desired to measure the angle α made by lines $P'S_1$ and $P'S_3$ and the angle β made by lines $P'S_3$ and $P'S_2$ where P' is the projection on the horizontal plane of the vehicle position P and S_1, S_2, S_3 are source positions. This may be accomplished by using a detector-collimator package as shown in Figure 11. If the rotation axis of the package is aligned with local vertical then the angle shown as α is measured as the angle made by rotating the collimator from $P'S_1$ to $P'S_3$. Likewise, the angle β is measured as the angle made by rotating the collimator from $P'S_3$ to $P'S_2$.

It is also required to measure the angle γ' made by lines PQ' and $P'Q'$ where Q' is the intersection of a perpendicular to the Y-axis from P' and a line parallel to the Y-axis drawn through S_3. This is accomplished by using a detector-collimator package as shown in Figure 11, aligning the rotation axis parallel to the Y-axis of the ground reference system. The angle γ' is the same as the angle which PQ' makes with the horizontal plane established by the vehicle inertial system. The angle γ' is then measured as the angle of rotation from the horizontal plane established by the inertial reference system to the line PQ'.

Measurement of the angle α establishes a circle on the ground plane which includes points P', S_1 and S_3 and has a radius $R_\alpha = \overline{S_1S_3}/2\tan\alpha$ (where $\overline{S_1S_3}$ denotes the distance between points S_1 and S_3). Likewise, measurement of the angle β establishes a circle on the ground plane which includes points P', S_2 and S_3 and has a radius $R_\beta = \overline{S_2S_3}/2\tan\beta$. These two circles intersect in two points only, S_3 and P', hence the X, Y coordinates of P' are established by measurement of the angles α and β . Having established P' we know the length of line $P'Q'$, since the separation and angular relation of S_1, S_2 and S_3 is fixed by ground construction. Measurement of the angle γ' then permits calculation of the length of line PP' from $PP' = P'Q'\tan\gamma'$. The X-Y-Z coordinates of point P' have been established, thus the vector $\overline{PO}$ and the position of the vehicle can be determined where O is the desired point of touchdown.

The following is a discussion of technique 2.2.(a-2), Ground Plane and Line of Sources Plane Angle Measurements (Figure 5). It is desired to measure the angles α and β as shown in Figure 5. These measurements proceed as for the previous case, technique 2.2.(a-1) and again the X-Y coordinates of point P' are established. It is also required to measure the angle γ formed by lines S_1P'' and S_2P'' where P'' is the projection of point P on the Y-Z plane of the ground reference system. This may be accomplished by aligning the rotation axis of a detector-collimator package as shown in Figure 11, parallel to the X-axis of the ground reference system. The angle γ is then measured as the angle of rotation made by the collimator's rotating from line $P''S_2$ to line $P''S_1$. Measurement of the angle γ establishes a circle in the Y-Z plane which includes P'', S_1 and S_2. Since the Y coordinate of P has been found by the measurement of the angles α and β, and is the same as the Y coordinate of P'', the Z coordinate of P'' is also established. Thus the X-Y-Z coordinates of point P have been established, allowing the vector $\overline{PO}$ and the position of the vehicle to be determined.

The final system requiring the use of rotating collimators is technique 2.2(a-3), Reference Plane Angle Measurements. It is required to measure the angles α, β and γ as shown in Figure 6. The measurements proceed as before except that in this instance it is necessary to use three detector-collimator packages, one with its rotation axis parallel to the Z axis of the ground reference system to measure α, another with its rotation axis parallel to the X axis to measure β and the other with its rotation axis parallel to the Y axis

to measure γ. Measurement of these angles then establishes circles in each of the three planes, the equations of which are involved with two of the three coordinates of point P. Solution of the three simultaneous equations yields the X-Y-Z coordinates of point P, and hence the determination of the vector $\overline{PO}$ and the position of the vehicle.

b. Problems - All of the rotating collimator systems depend on angular measurements to determine the vehicle position. The accuracy required for these measurements depends on:

(1) The desired position accuracy.

(2) The accuracy of the vehicle inertial system.

(3) Accuracy of the servo system used to stabilize the platform upon which the detector-collimator packages are located.

(4) The geometrical arrangement of the ground-based radioactive sources with respect to the vehicle landing site.

Taking all of these into account it is found that the system certainly must be capable of measuring angles to within one degree and most likely less. For a measurement accuracy of one degree, the following must be considered:

(1) The location of the source must be determined as the collimator rotates past it. This limits the rotation speed for reasonably sized sources and detector-collimator packages.

(2) The detector-collimator package size and weight must be held to a minimum. This imposes a lower limit on the source size for a given configuration.

(3) The information received at the detector is statistical. This tends to cause an increase in source strength and/or detector size.

(4) The information rate from the detector-collimator package must be of the order of five bits per second. This causes either an increase in rotation speed or an increase in detector-collimator size and weight.

(5) The radiation field must be sampled in discrete time units. This limits the rotation speed or causes an increase in source strength.

The design of the collimator-detector package places an intrinsic limit on the accuracy of the angular measurement. Since a finite sampling period is required it is necessary to record a statistically significant number of counts during the period that the detector sees the source. In general, one may determine within one period either side of the true period when one locates the source. The accuracy requirement on the collimator is thus increased to one third of one degree. This requires that the collimator vanes be three hundred and forty-four times as long as the spaces between them. For a six inch collimator, corresponding to an eighteen inch diameter package, this spacing is only 0.0175 inches. The thickness of the vanes must then be less than 0.01 inches if a reasonable portion of the detector area is to be utilized.

3.2 SERVOED SYSTEMS

a. Concept - As in the case of the rotating collimator systems these systems also depend on measuring angles to determine the position

of the vehicle. In these cases, however, each detector package continually looks at one source and thus the system is not required to make rapid determinations of locations. It is also possible to do away with the need for alignment with true north, retaining only the establishment of local vertical, and hence a horizontal plane, by the vehicle inertial system. However, one adds a complication to the calculation by not measuring the angles in the mutually perpendicular planes of the ground reference system.

It is necessary to measure the angles α and β as shown in Figure 9. This may be accomplished by using either a collimator with horizontal and vertical elements, denoted here as the Egg Carton Collimator as shown in Figure 7, or a Blinded Detector System as shown in Figure 8. In either case the package is servoed both in azimuth and elevation to yield the maximum count rate. For the Egg Carton Collimator this is the only criterion for determining the direction of the sources. In the case of the Blinded Detector System, however, the measurement is augmented by requiring only that equal count rates be seen in each of the three detectors of a package. In both instances one detector package is required to track each source. In the case of the Egg Carton Collimator the packages are identical; however, for the Blinded Detector System the left hand package is the mirror image of the right hand package. It is seen that a total of two detectors is required in one instance, while six detectors are required in the other. In either case once the packages have been locked onto the two sources and a local vertical established the angles α and β may be read from

resolvers. It is also required to measure the angle γ, the angle between lines PS_1 and PS_2. This is easily accomplished since γ is the included angle between the two detector packages.

The measurement of α and β determines two concentric right circular cones, each of which includes one of the sources S_1 or S_2 in its intersection with the ground plane. Since the height of each cone is the same, the ratio of the radii of the two circles formed by the intersection of the two cones with the ground plane (X-Y plane) may be determined from $R_\beta / R_\alpha = \tan \beta / \tan \alpha$. The locus of points of the intersection of two lines, the ratio of whose distances from two distinct points is a constant, is a sphere. This sphere intersects the ground plane in a circle. Measurement of α, β and γ and the establishment of the local vertical and hence a horizontal plane allows rotation of γ from the measurement plane to the horizontal plane where it has a value γ'. The angle γ' is also the ground plane angle between $P'S_1$ and $P'S_2$. The locus of points of the vertex of a given angle whose sides include S_1 and S_2 respectively is also a circle in the ground plane. The intersection of the two ground plane circles establishes the X-Y coordinates of point P'. The Z component of point P may now be found from using the X-Y components of P' and either the angle α or the angle β. The vector $\overline{PO}$ and the position of the vehicle is thus determined. It must be pointed out that there is some ambiguity in the location of point P'; however, it is not critical since the ambiguous point lies on the other side of the line S_1S_2 and may be discredited if track is known to within 180°.

b. Problems - As for the rotating collimator systems, the servoed systems depend on angular measurements to determine the vehicle position. The accuracy required for these measurements depends on the following factors:

(1) The desired position accuracy.

(2) The accuracy of the vehicle inertial system.

(3) The accuracy of the azimuth and elevation servo systems.

(4) The geometrical arrangement of the ground based radioactive sources with respect to the vehicle landing site.

Again the accuracy requirement is certainly within one degree and most likely better. The Egg Carton Collimator has the same problem attributed to the rotating collimator, i.e., many closely spaced vanes. In addition the intrinsic accuracy must be better since the detector must continuously search, which introduces an additional error not found in the rotating collimator systems. This searching is required until the error signal driving the servo system goes to zero. In this case the error signal is detection of less than maximum intensity which may only be verified by searching for a greater intensity. The Egg Carton Collimator has a more serious difficulty in that the acceptance angle for signals is very small making acquisition difficult. This may only remedied by incorporation of additional detectors.

In the case of the Blinded Detector System the intrinsic accuracy is limited only by the size of the detector-blinder package. There is however, another problem which crops up since relative intensity measurements are used to position the package. The positioning

then may only be performed to within the accuracy established by the statistics of the counts coming from the sources. This effect establishes a random motion of the package which is more severe the lower the count rate. The most serious difficulty arises from the necessity of incorporating six detectors, NaI(Tl) scintillation crystals, in the system. These crystal-phototube assemblies must be matched if the same count rate for exposure to a fixed intensity in each is to be realized. This problem is made more severe by the necessity for using pulse height discrimination to bias out system noise.

3.3 MATRIX SYSTEMS

a. Concept - This technique depends on the mapping of some region of space by means of ground collimated radiation patterns such as are shown in Figure 12. This pattern is feasible for conventional type aircraft. For VTOL, a more complicated type of matrix incorporating altitude information would be formed above a prescribed landing area. Position information would have to be determined by measuring relative intensity or counting the presence or absence of radiation. One advantage of this type of system is that only one detector and no stabilization of the detector package is required.

b. Problems - In any of these types of methods there are serious problems involved. In order to produce the required matrix, intricate collimation is required along with relatively high source strengths. It is also necessary to recognize the pattern at the aircraft and relate it to a known pattern to obtain position. As pointed out earlier, it requires 5000 cps to identify a well collimated radiation area

in one millisecond. Traveling through a matrix requires rapid recognition and a sophisticated bookkeeping system to determine position, provided the beams are well defined. At large distances where build-up becomes a factor, the region between the beams becomes quite fuzzy and the pattern less defined. Also, the start of the pattern must be determined either through energy or intensity measurements which complicates the source detection system, and bookkeeping of the system. For these reasons, the matrix system does not appear to be practical.

3.4 INTENSITY

a. Absolute Intensity

(1) Concept - In this scheme, as shown in Figure 10, the absolute intensity from three sources is measured. From these measurements the distances r_1, r_2 and r_3 may be determined. From two sources the distances r_1 and r_2 may be determined. These distances together with the included angle γ gives the X-Y-Z coordinates of point P. This point can be obtained by determining r_1 and r_3 whereupon knowing the distance between S_1 and S_3, it is possible to determine the angle α. Similarly, it is possible to measure the angle β. Since the point P lies on circles determined by r_1/r_3 = constant and r_2/r_3 = constant, the intersection of these two circles is the point P above the X-Y plane.

(2) Problems - In order for this system to work, it is necessary that all three detectors be closely matched, electronics of very good stability, especially the gain stability of the photo tubes must be good. Calibrations must be made to account for source decay,

variations in attenuation in the air and source position errors. Keeping all the parameters constant is an impossible job. This is further complicated by the fact that each aircraft must have a calibration for each different landing site. The biggest problem of all is that the measurement requires good statistics meaning long counting time or high intensity sources. Because of the stability, calibration, and excellent statistics required the absolute intensity measurement systems do not show promise.

b. Relative Intensity

(1) Concept - In this method, the outputs of two detectors are compared, producing an error signal which allows the vehicle to be directed to points equidistant from two sources. It is necessary to provide another measurement since this only provides localizer information. It is to be noted that most of the calibration problems associated with absolute intensity measurement are avoided by this technique.

(2) Problems - Since the ratio of the outputs of the detectors is used, it is necessary that the detector packages be closely matched. In order to obtain signals suitable for comparison, the statistics must be good or larger measurement errors must be accepted. A further requirement on the statistics is imposed if relatively fast response is desired, since the time constant associated with the count rate determination must be short. A short time constant implies high count rates for accurate measurement. The main disadvantage of this system is the problem of statistics.

3.5 ENERGY DISCRIMINATION

a. Concept - The unique characteristic gamma rays of radioisotopes can be used to differentiate different sources or to provide a means to signal the start or end of a certain measurement as required in the matrix method. The spectral region of maximum count rate per unit energy is detected and the energy determined using a pulse height analyzer (PHA). The PHA window is set for the required energy and only those gamma rays which satisfy the PHA requirements are counted, all others are discarded.

b. Problems - The need for a narrow window to select a spectral region requires that the source strength be high. Only limited information can be obtained using this method. It must be used in conjunction with other techniques to determine position.

3.6 INTENSITY MODULATION

a. Concept - In this method a shutter over a lead well is used to modulate the flux of gamma rays out of the shield. By sensing the modulation of the source, it is possible to identify the radiation source. Detection of this pulsed beam cannot of itself give position information.

b. Problems - The pulse frequency of a shuttered well will be low since a mechanical shutter system is necessary. This implies that the rate of information will be low. In order to obtain additional information, greater repetition rates must be used implying increased source strength.

4.0 CONCLUSIONS

Based on the preceding review of radioisotope techniques for the acquisition of Control Data during the terminal phase of VTOL flights, the following conclusions are drawn:

a. The greatest promise for the application of radioisotope systems appears to be in the terminal phase, i.e., range from 600 feet to touchdown.

b. The rotating collimator systems have problems which require additional investigation before a final assessment of their practicality may be made. These problems include development of parameters related to the trade-offs among large detector-collimator packages, high intensity sources, and information rate and the optimization of the airborne stabilization system, detector-collimator package, and associated data processing elements.

c. A servoed detector system, i.e., the blinded detector system, seems to be feasible, however, since no experience is available in the airborne application of this system, certain experimental investigations are required. These include determination of speed of response, ultimate angular resolution, package size and weight, and data processing.

d. Since matrix systems involve the use of intricate ground based collimation, implying precise alignment, and complicated on board data processing which includes pattern recognition and a sophisticated bookkeeping system, their application to this problem does not appear to be practical.

e. The absolute intensity system is plagued by problems of stability of electronics, calibration, matched electronics, and excellent statistics required which implies large source strengths. For these reasons, systems involved in the measurement of absolute intensity do not show promise.

f. Relative intensity, energy discrimination, and intensity modulation measurement systems may only be used to augment other systems. It is impossible to use these systems by themselves to determine vehicle position.

g. The problems of roll, pitch, and yaw may be overcome by the use of an on-board inertial system to either align the detector-collimator packages or to supply appropriate angles for correction of the measured data.

h. A major factor in determining source strength requirements is the build-up factor. Better experimental data on build-up factors is necessary.

i. While the application of radioisotope techniques to the terminal phase of VTOL landing has not yielded a clear-cut solution, this does not preclude their successful application to other control problems. This is particularly true where the distances involved are not large or where disturbing media such as large thicknesses of atmosphere are not present. Although several ideas have been discussed here, of which only two show promise of application with additional investigation, it is not to be inferred that other solutions involving radioisotope techniques may not be conceived.

5.0 RECOMMENDATIONS:

The word recommendations is used here only for want of a better word. Because of the potentials which may be realized, it is to the advantage of the USAF to continue the investigation of these techniques. Specifically, the following program must be initiated.

a. Further investigation into the feasibility of using the rotating collimator system is required. Specifically, it is necessary to (1) develop parameters related to trade-offs among large detector-collimator packages high intensity sources, and information rate; (2) perform detailed design and optimization of the airborne stabilization system and the detector-collimator package; (3) develop appropriate data processing elements.

b. Further investigation into the feasibility of using servoed blinded detector systems is required. Specifically, it is necessary to (1) determine the practical speeds of response; (2) determine the ultimate angular resolution of the detector package; (3) optimize the detector package and associated servo components for size and weight; (4) develop appropriate data processing elements.

c. Initiate experimental program to determine build-up factor as a function of energy, collimation, and distance.

REFERENCES

1. Private Communication, Erkinif, N., SESJ, Research and Technology Division, Wright-Patterson AFB, Ohio

2. "Nucleonic Aircraft Landing Systems" Industrial Nucleonic Corporation, Columbus, Ohio, April, 1964

3. Beavin, R. C., 1st Lt., "Radiation Physics - Its Impact on Instrumentation" ASD TDR 63-697, September, 1963

4. Haefner, K. B., et al, "Radioisotopes for Low Altitude Measurement", ASD TDR 62-648, July, 1962

5. Meiselman, J., Beavin, R. C., 1st Lt., Polishuk, P., "Tactical Landing System" FDCL-TM-63-28, November 1963

6. Private Communication, J. N. Maddox, AEC Division of Isotope Development, Germantown, Maryland, January 1964

7. Davis, J. P., Reinhardt, P. W., "Instrumentation in Aircraft for Radiation Measurements" Nuclear Science and Engineering 2, 713 - 727 (1957)

BIBLIOGRAPHY

1. "Tactical Army Aircraft Landing Systems" (TAALS) Second Quarterly Progress Report, 1 November 1963 - 31 January 1964 DA-36-039-AMC-03367(E)

2. Innis, R. C., Quigley, H. C., "A Flight Examination of Operating Problems of V/STOL Aircraft in STOL type Landing and Approach" NASA TM D-862, June 1961

3. "Preliminary Requirements Study for V/STOL", Lear-Siegler, Inc. Progress Reports 1 - 4 AF33(657)-11740

4. "State of the Art for V/STOL Control Display" RTD-TDR-63-4167 October 1963, Lear-Siegler Corp., AF33(657)-11740

5. Bird, L. L., Ziegler, C. A., "Low Level Radiation Altimeter System Study" Parametrics, Inc., NAS9-1490, October 1963

6. Campbell, J. P., "Vertical Takeoff and Landing Aircraft" The MacMillan Company, New York (1962)

7. "Approaching Transonic Commercial Air Travel Propels All-Weather Landing Systems to the Fore" IEEE Spectrum, Vol 1, No. 1, P177 January 1964

8. "Vertical Take Off and Landing Aircraft" Annals of the New York Academy of Sciences Vol 107 Art 1, March 1963

9. Anderson, S., "An Examination of Handling Qualities Criteria for V/STOL Aircraft" NASA TMD-331, July 1960

10. Gerles, R. M., Weick, R. F., "A Preliminary Piloted Simulator and Flight Study of Weight Control Requirements for VTOL Aircraft" NASA TMD-1201, February 1962

11. Garren, J. F., Jr., A. Assadourian "VTOL Weight Control Requirements in Hovering as Determined from Motion Simulator Study" NASA TND-1488, October 1962

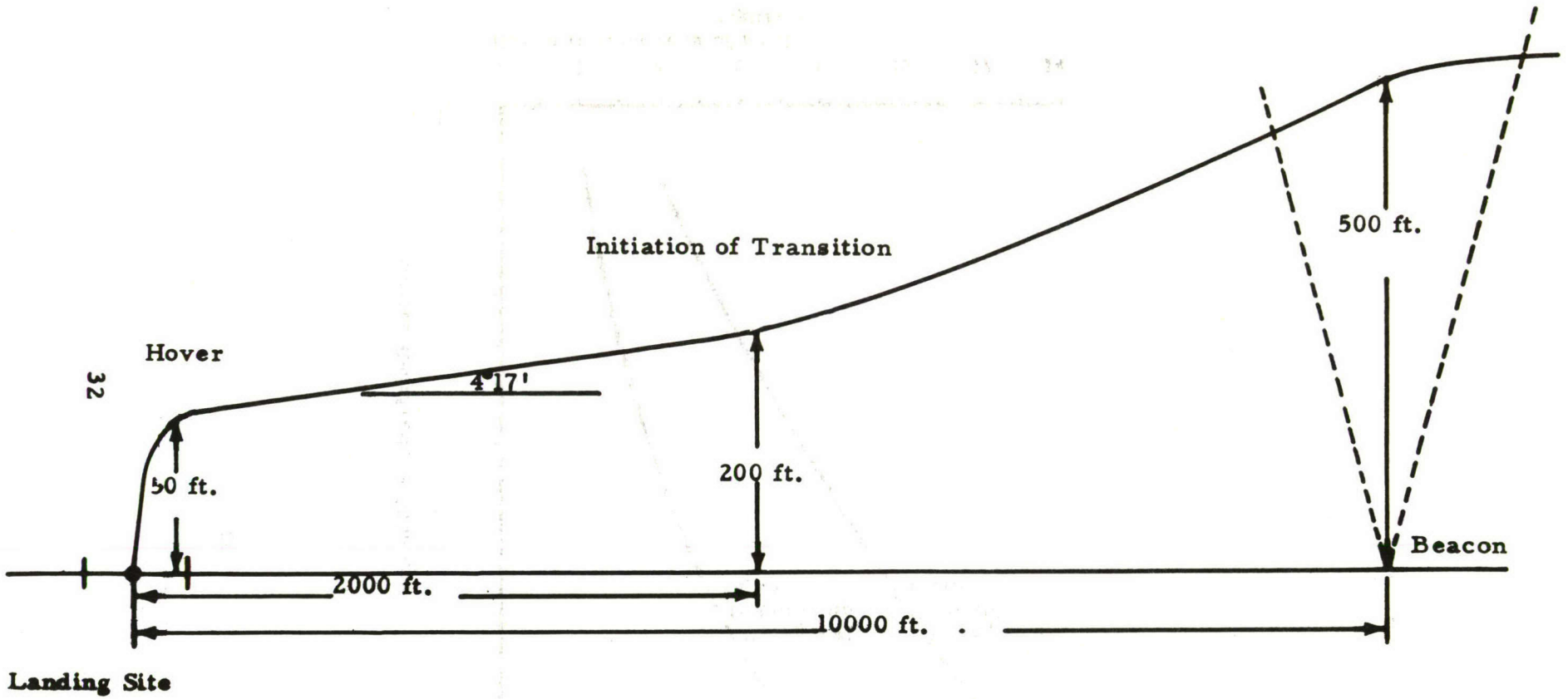
12. Sullivan, M., "Control of VTOL Aircraft" Sperry Engineering Review, Winter, 1963

13. Anderson, S. B., "An Examination of Handling Qualities Criteria for V/STOL Aircraft" NASA TMD-331, July 1960

14. Foody, J. F., "Controlling VTOL Aircraft at Transition Speeds" Control Vol 3, pp 88-91, September 1960

15. Ibid (November, 1960) pp 107-111

16. Isca, J. A., Patierno, J. "Instrument Flight Simulator Study of the VTOL Controllability - Control Power Relationship" IAS Paper 61-118-1812, June 1961



TYPICAL LANDING PROFILE FOR TACTICAL VTOL

Figure 1

REQUIRED SOURCE STRENGTHS AS A FUNCTION OF RANGE

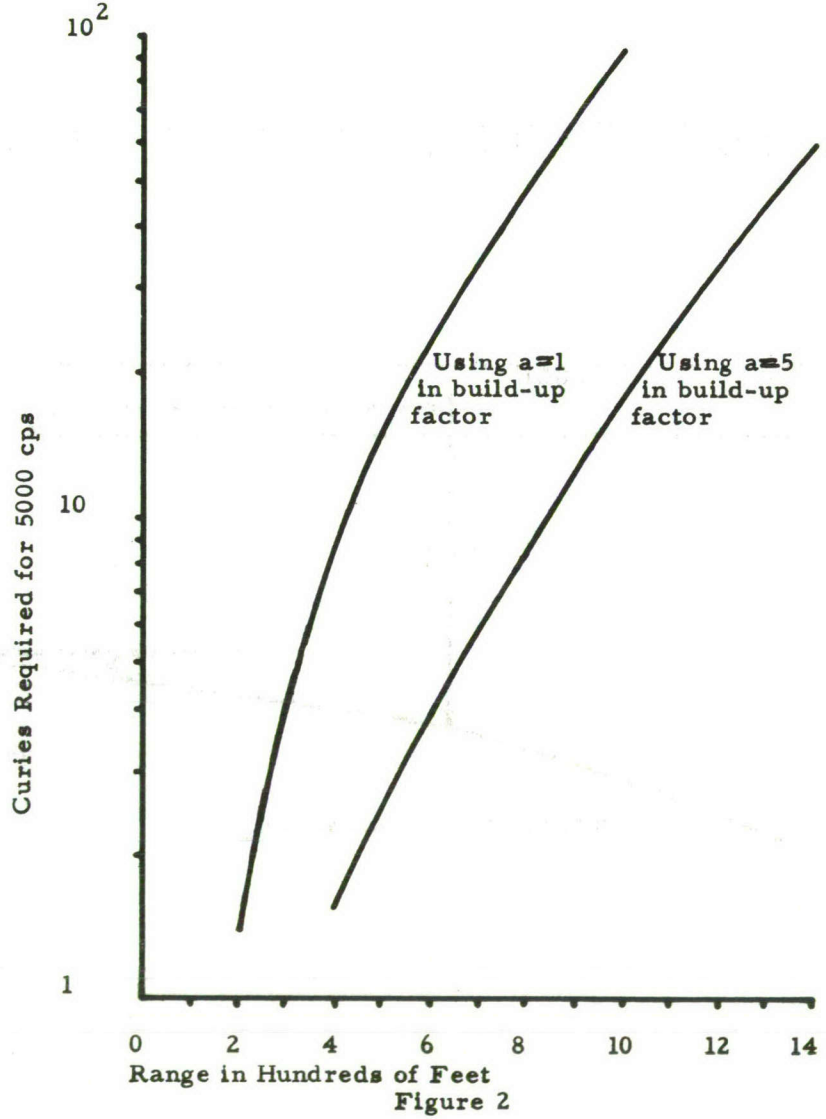
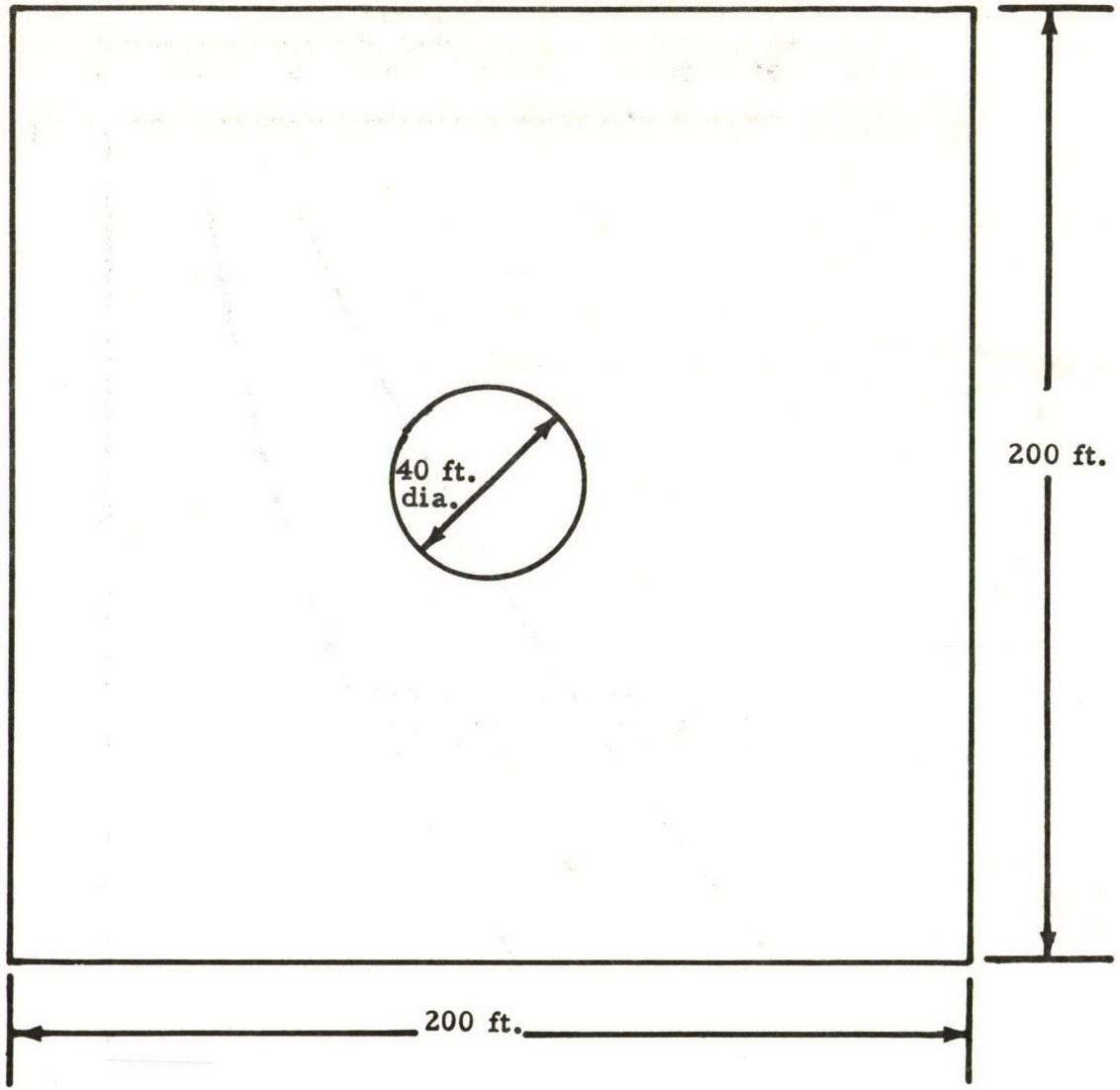
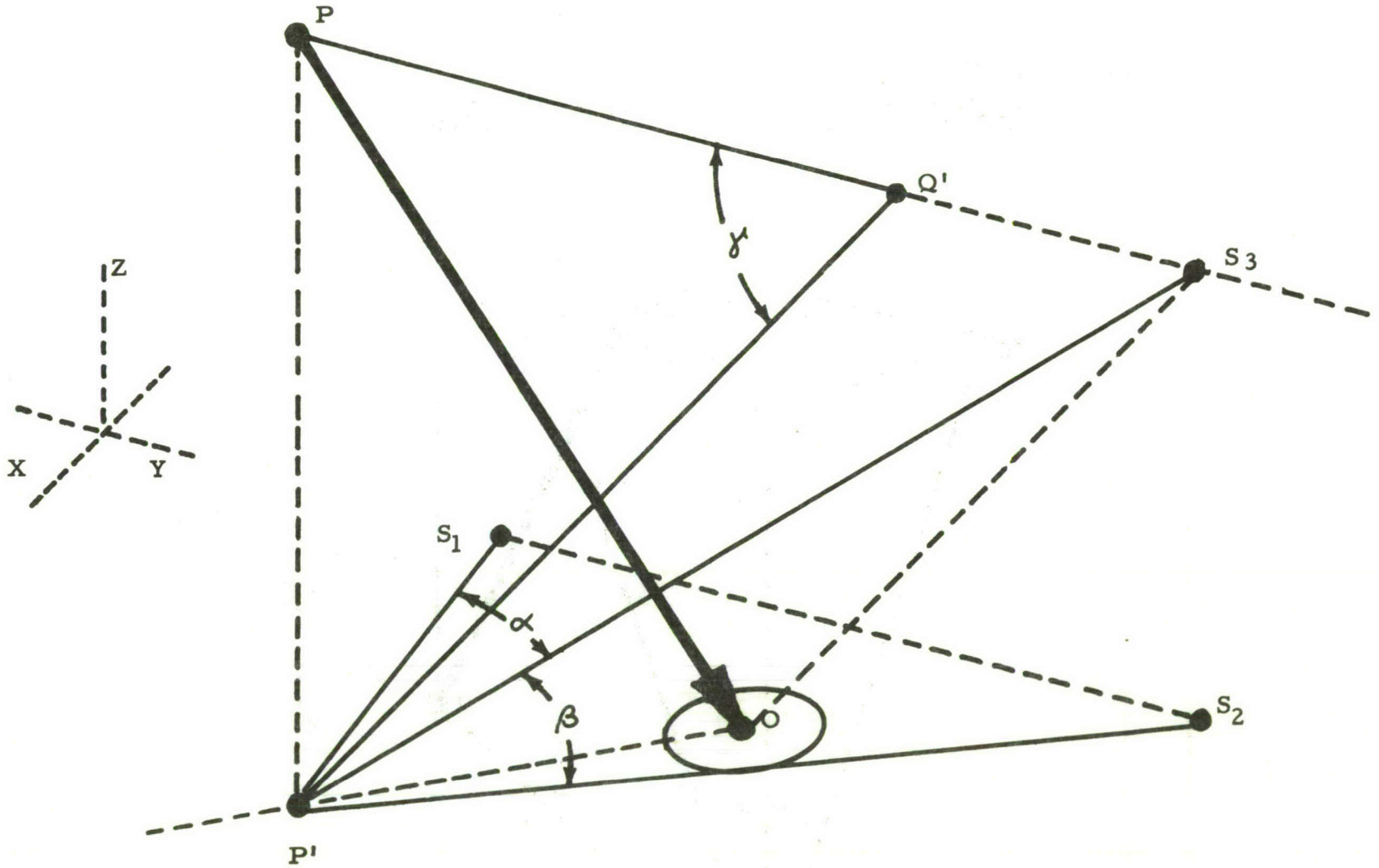


Figure 2



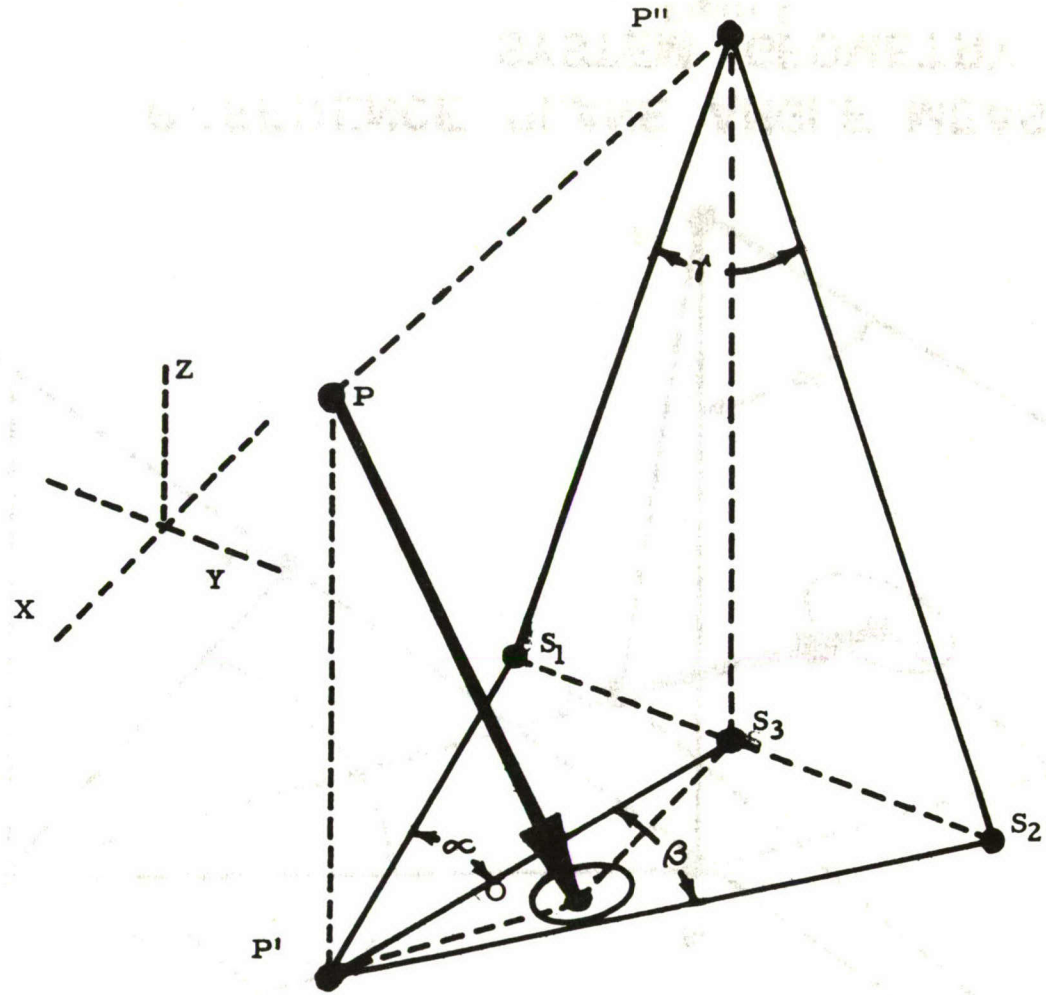
FORWARD VTOL SITE

Figure 3



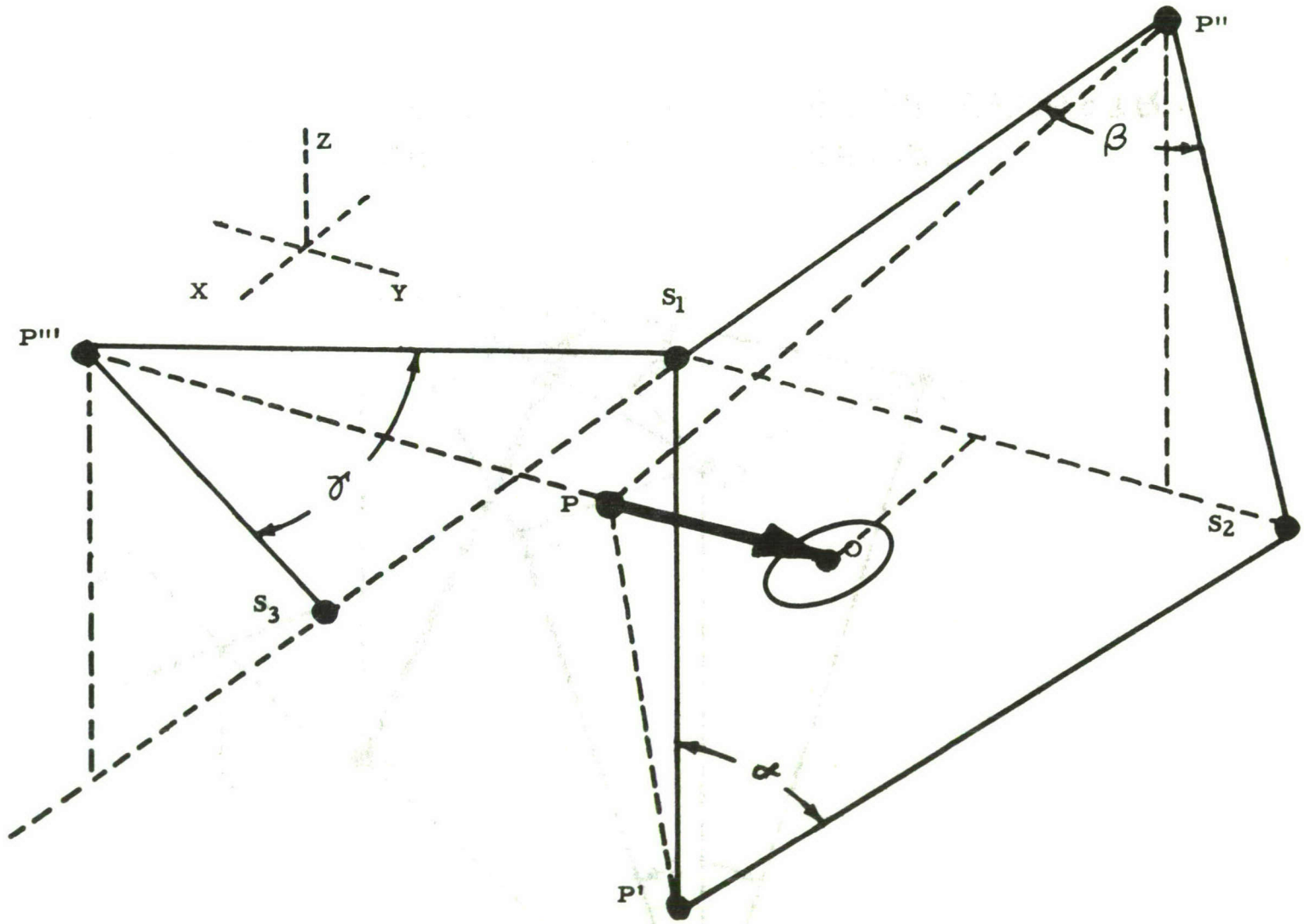
GROUND PLANE AND ELEVATION ANGLE MEASUREMENT SYSTEM GEOMETRY

Figure 4



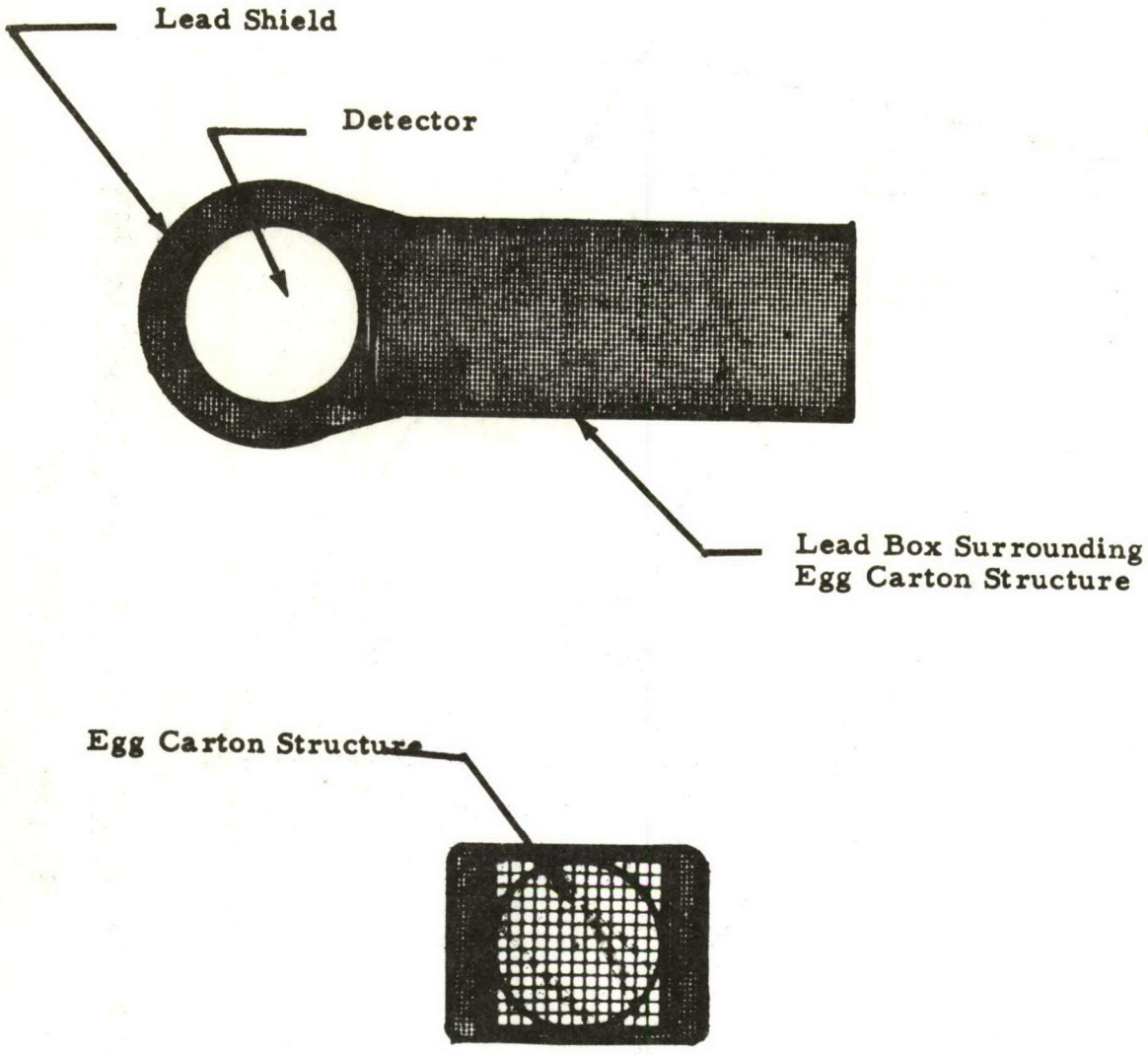
GROUND PLANE AND LINE OF SOURCES
PLANE ANGLE MEASUREMENT SYSTEM GEOMETRY

Figure 5



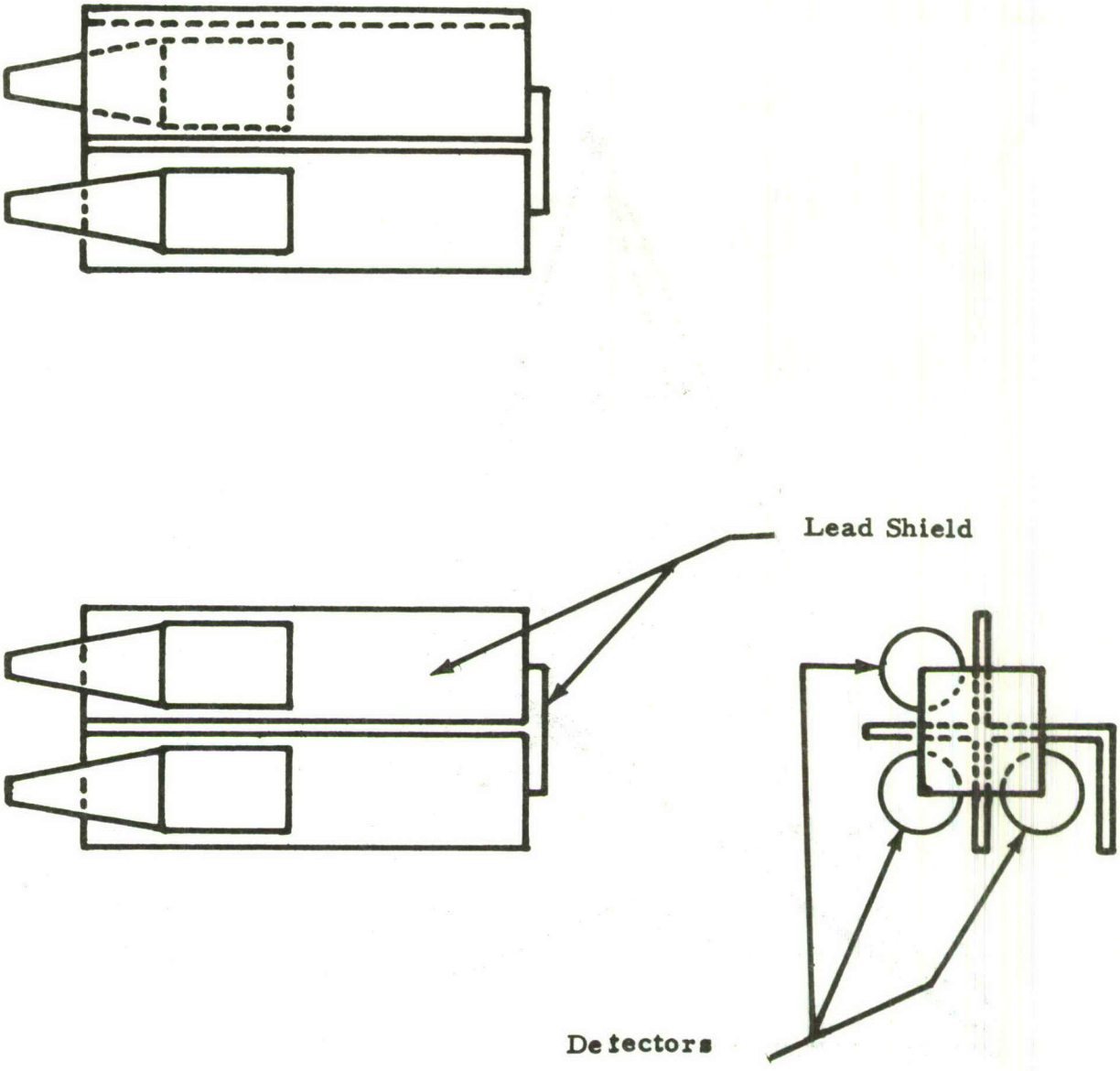
REFERENCE PLANE ANGLE MEASUREMENT SYSTEM GEOMETRY

Figure 6



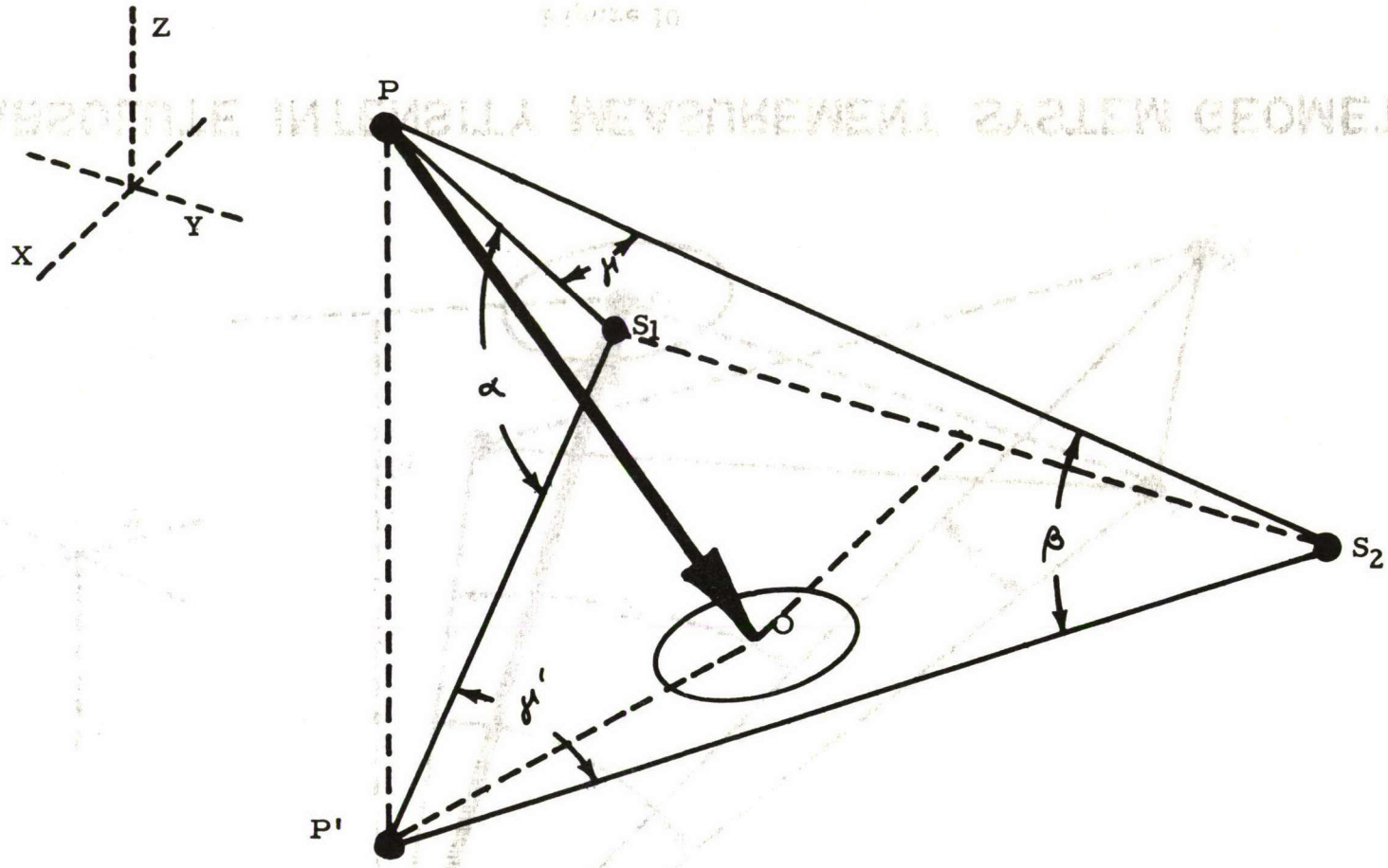
EGG CARTON COLLIMATOR-DETECTOR PACKAGE

Figure 7



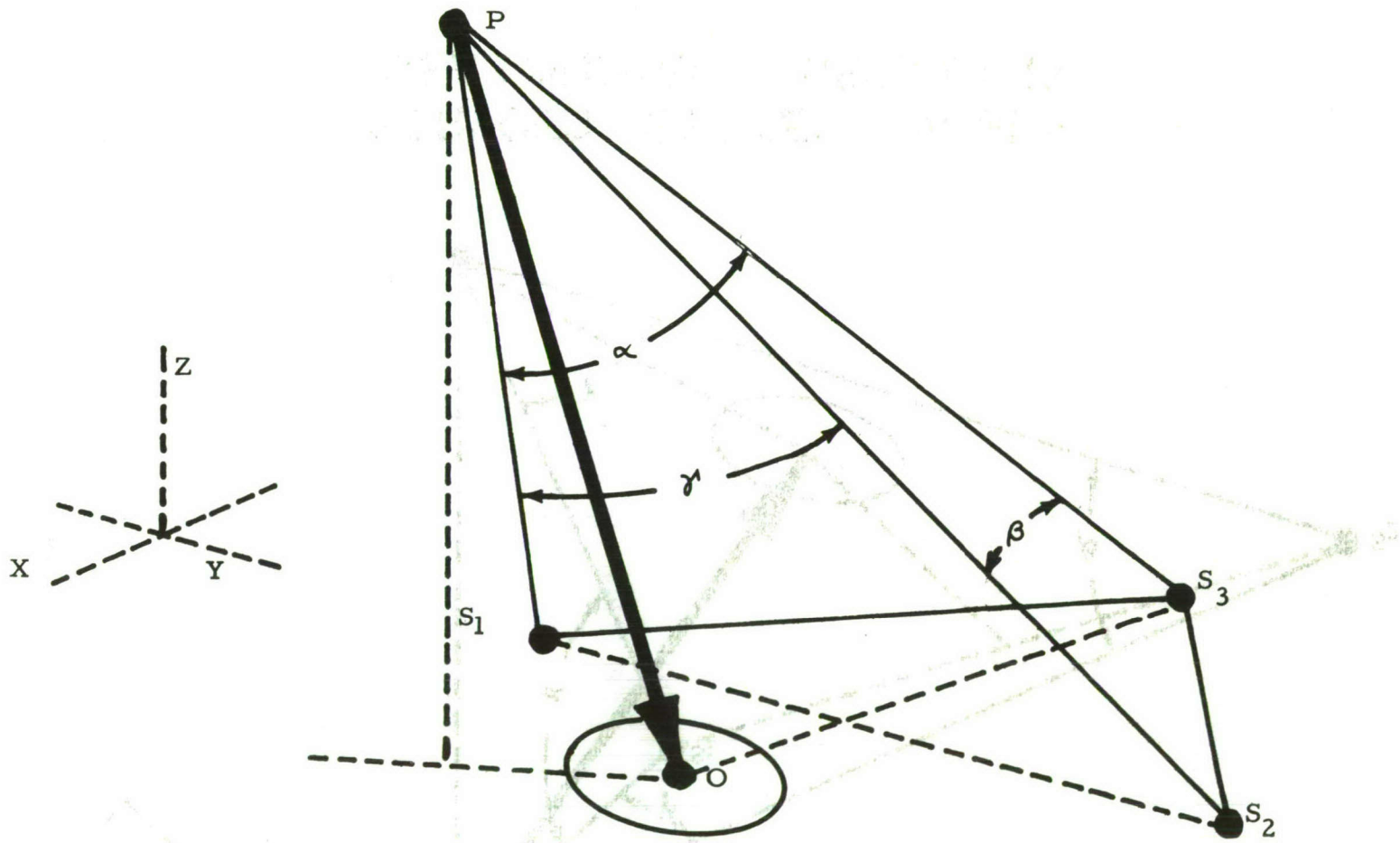
BLINDED DETECTOR-COLLIMATOR PACKAGE (Ref. 2)

Figure 8



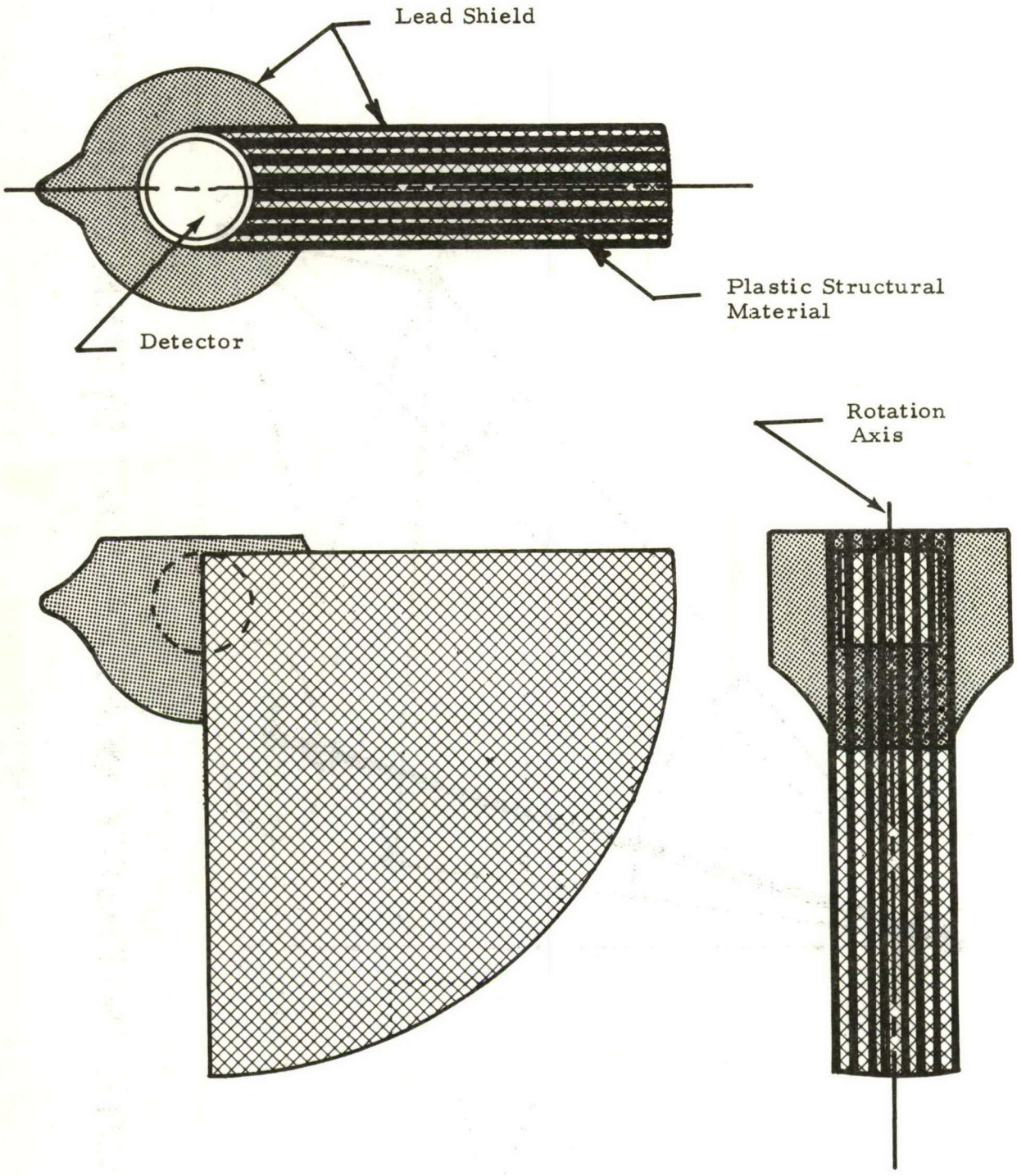
SERVOED SYSTEM ANGLE MEASUREMENT GEOMETRY

Figure 9



ABSOLUTE INTENSITY MEASUREMENT SYSTEM GEOMETRY

Figure 10



ROTATING COLLIMATOR-DETECTOR PACKAGE

Figure 11

MATRIX PATTERN FOR CONVENTIONAL A/C

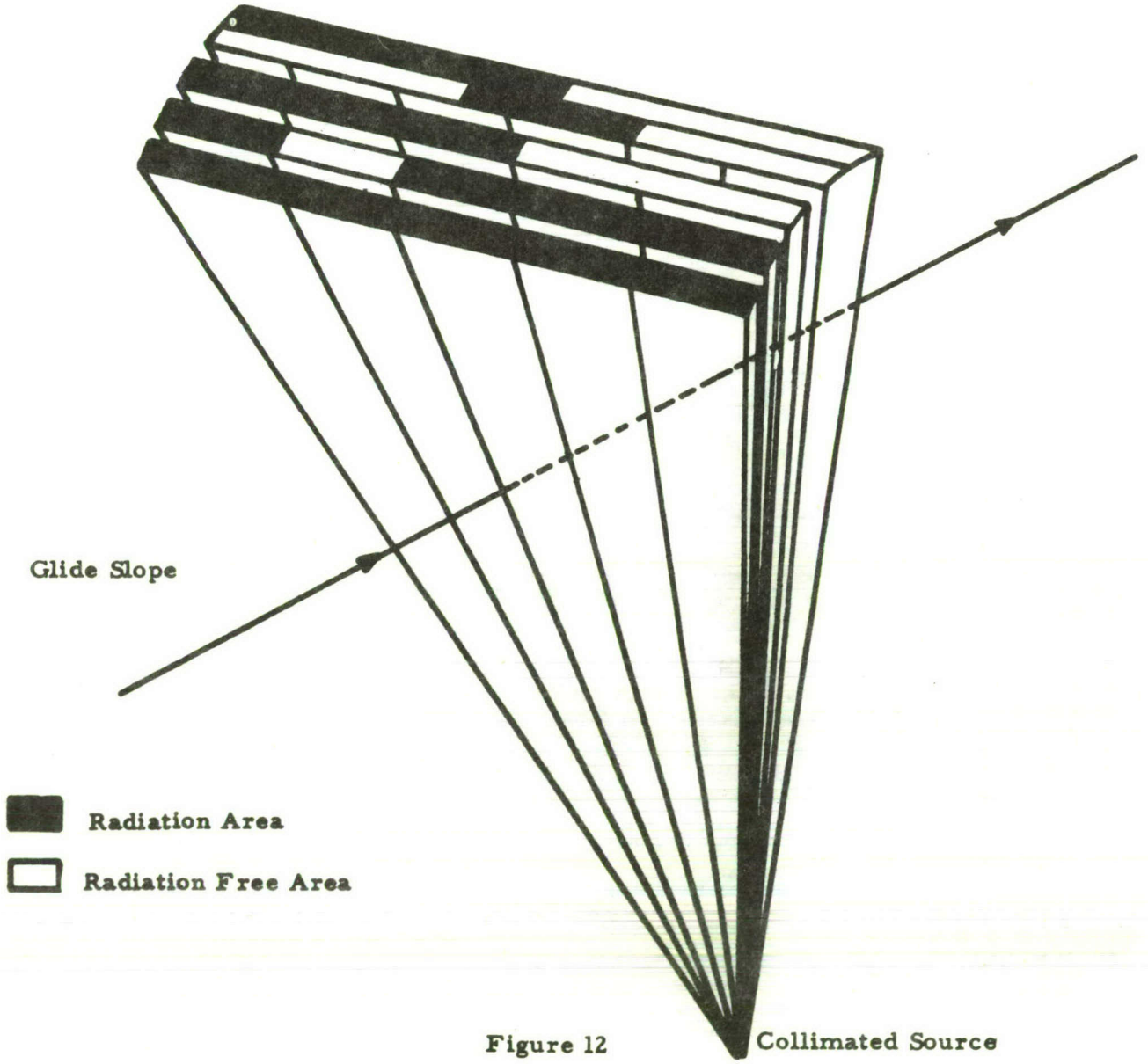


Figure 12

AFPS SA OCT 64, 500